228

R-582 Rev I

ASSEMBLY OF COMPUTERS TO COMMAND AND CONTROL A ROBOT

by
Louis L. Sutro
and
William L. Kilmer

February 1969

Revised December 1969

INSTRUMENTATION LABORATORY

07641

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Cambridge 39, Mass.

R-582 Rev I
ASSEMBLY OF COMPUTERS TO COMMAND AND CONTROL A ROBOT

by

Louis L. Sutro

Assistant Director
Instrumentation Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

and

William L. Kilmer

Assoc. Professor of Electrical Engineering and System Science
Michigan State University
East Lansing, Michigan

February 1969
Revised December 1969

Presented at the
Spring Joint Computer Conference
Boston, Massachusetts, May 1969,
and printed in the Proceedings of
that conference, AFIPS Press,
Montvale, New Jersey.

INSTRUMENTATION LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

Approved: Yao Tzu Li
Yao T. Li
Professor of Aeronautical
and Astronautical Engineering
Massachusetts Institute of Technology

Approved: John W. Hursh
John W. Hursh
Associate Director
MIT Instrumentation Laboratory

Approved: R B Woodbury
R. B. Woodbury
Deputy Director
MIT Instrumentation Laboratory

ACKNOWLEDGMENT

Because the physiological concepts presented here came from Warren McCulloch it would have been correct to list him as an author, but in doing so we could not have identified his contributions. One who helped us describe these contributions was Dr. Jerome Y. Lettvin. The simulation of the visual first-stage computer is the work of James Bever who began it, John Hatfield who devised the form of filter employed in level 5, and Jerome Lerman who devised all of the filters presently employed. The two simulations of retic were carried out by Jay Blum. John Leavitt designed the hardware version of the visual first-stage computer.

Four agencies of the U.S. Government made possible the work reported here. Research into animal vision, development of visual computers and programing of S-RETIC were supported by the Bioscience Programs of NASA's Office of Space Science and Applications.* The Air Force Office of Scientific Research** supported the research into retic and the analysis on which S-RETIC and STC-RETIC were based. The NASA Electronics Research Center*** supported the programing and initial testing of STC-RETIC and the assembling of subsystem concepts into a system (robot) concept. The U.S. Air Force bionics programs**** supported the extended testing of STC-RETIC and the design of H-RETIC.

The publication of this report does not constitute approval by the sponsoring agencies of the findings or the conclusions contained therein. It is published only for the exchange and stimulation of ideas.

* Contract NSR 22-009-138 through Instrumentation Laboratory, MIT.
** Grant AF-AFOSR-1023-66 through Michigan State University.
*** Grant NGR 22-009-140 through Instrumentation Laboratory, MIT.
**** Contract AF 33(615)-3885 through Instrumentation Laboratory, MIT.

R-582 Rev I

ASSEMBLY OF COMPUTERS TO COMMAND AND CONTROL A ROBOT

ABSTRACT

Robot behavior, previously demonstrated only by higher animals, especially man, is being made possible by a description of the vertebrate nervous system in engineering terms and by the design of computers fulfilling these descriptions. Like man, a robot requires computation in five major domains: (1) Sensory (predominantly visual), (2) command, (3) relational, (4) timing, coordinating and auto-correlating and (5) effector sequences.

Visual computers, simulated in a general-purpose computer, have enhanced the contrast of a stereo pair of images, formed a stereo pair of line drawings, and computed the range of details. A command computer, also in simulation, has selected the appropriate mode of behavior, which determines the filters used in the visual computers. Moreover, the simulated command computer has stored a reduced record of the conditions under which it chose each mode, a record which corresponds to development and conditioning in animals. Relational computers and computers of effector sequences are being simulated elsewhere. A visual computer is under construction in hardware. A command computer is being designed.

by Louis L. Sutro
and
William L. Kilmer

February 1969

Revised December 1969

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

I. INTRODUCTION

There is a growing consensus among predictors of science that the world is about to witness the evolution of what might be called a new species -- the robot. Whereas, animal evolution was a trial-and-error process, robot evolution appears likely to be carefully contrived. Starting where animal evolution left off, that is, with man, robot evolution promises to excel man in some respects, and be excelled by him in others.

To the computer profession, one challenge in this progression is to develop computers for robots that match those that have been found indispensable in men. We are aided in this task by the description of the human nervous system in computer terms by physiologists such as Warren McCulloch.

With his description before us, we have devised working models of two of the five principal computational domains which he identifies in the nervous system of vertebrates, including man. Others are devising working models of other domains. Implemented in light, portable hardware and connected together, these computers promise to provide intelligence for a system that will sense its environment, move about and perform useful tasks.

Who needs a robot? Everyone who would like help with tiring chores. However, early models with large arms and wide wheelbases cannot move around the home or office. One need that has led to the development about to be described is exploration of the planet Mars. For this task, robot development is being pursued not as an end in itself but as a framework within which to develop an automatic visual subsystem. A second need is for a computer to command a system receiving several forms of input, such as sight, sound, touch, and reports on its own movements. Here again robot development provides the framework for the computer development.

As well as can be determined,(1) the surface of Mars is open country where a wide-wheelbase vehicle should be at home. More to the point, the only exploration there for a decade or more will have to be either by a remote-controlled or an automatic vehicle. The distance is such that a single bit of information requires 15 minutes, on the average, for transmission from Mars to earth. With such a transmission delay, remote control seems hardly practical. An automatic vehicle or robot thus seems imperative.

While the surface of Mars is colder than the surface of the earth, there may be hot spots due to volcanic or other sub-surface activity. All the moisture on Mars, according to our instruments, is in the form of either gas or ice. The atmospheric pressure is too low to hold it as water, but it might pass through the water phase in these hot spots, lasting as water long enough to make possible life as we know it.[2]

To go to these hot spots, if indeed they exist, poke around them, pick up and examine samples seems the best way of finding out what is there. Even if there is no life on Mars, there are cliffs formed at the edges of craters, that need to be examined for their geology. The craters need to be climbed into and out of. To go from one crater to another, crossing must be made of the ravines called "canals".

2. RESEARCH AND DEVELOPMENT

The robot design described here began as an effort to design eyes for the artificial intelligence that Marvin Minsky and John McCarthy called our attention to, in the fall of 1958. Persuaded that eyes for artificial intelligence could be achieved only by employing ideas from animal vision, one of us (Sutro) approached Dr. McCulloch for advice. The colloboration that ensued led first to an analytical model of an animal retina that recognizes objects, namely, the retina of a frog[3, 4]. It led next to a proposal to NASA to develop means of reducing, for transmission to earth, pictorial data acquired in the search for evidence of both life and geological changes on Mars. Supported then by the NASA Bioscience Programs, we undertook this in the manner Dr. McCulloch and we thought best, namely, to model animal vision in lightweight, low-power hardware. Study of frog vision showed how recognition of a simple shape (a bug) can be achieved in two levels of computation, but it did not carry far enough the data reduction we felt was required. Needed, we felt, was reduction of a stereo pair of images on Mars to a pair of line drawings with shading, as we primates do. Geologists and biologists make line drawings with shading to represent what they see. The lines portray edges, angles and silhouettes. The shading conveys the brightness of surfaces.

Man forms in his head a model of what he observes. Formation of a line drawing with shadings is a stage in the computation of this model. However, as Dr. McCulloch points out, the vision of a primate cannot be modeled by itself. Data flows not only inward from the images, but outward from the brain to adjust the filters, focus, convergence and direction of gaze that select what will flow inward. For a visual system employed in a single position on Mars, these adjustments can be either preset or changed by commands from earth; but when the system is required to move about, the commands to adjust it can scarcely be sent from earth. They have to be generated on site.

To develop a command computer one of us (Kilmer) undertook to model the part of the vertebrate brain that decides from information received through all the senses what class of thing the animal will do from moment to moment. This is the core of the animal's reticular formation, extending through its brain stem and the length of its spinal cord. Support for its development came first and continues from the Air Force Office of Scientific Research, came then from NASA's Electronic Research Center, and comes now from the U. S. Air Force bionics programs.

Cameras and computers under development are pictured in Fig. 1. At the left is a binocular pair of TV cameras of which sufficient study has been made to indicate that each camera can be built to view the world through both wide- and narrow-angle lenses. Receiving the output of the camera is the visual first stage computer which enhances contrast in an image, as an animal retina does. Next to it are switching filtering and comparison structures, we call the visual second stage computers. A model of the environment consists of relations formed in this second-stage visual computer and stored in the visual part of the relational computer. A line, which indicates sharp change in luminance, is a relation of high spatial frequencies. Shading, which indicates the difference in luminance of areas, is a relation of low spatial frequencies. Each filter passes one band of frequencies more than others. Commands to adjust filters, focus and direction of gaze are shown as arrows rising from the command computer in Fig. 1. Since these commands will pass through structures not shown, the arrows are not drawn directly to the cameras and visual computers.

Note the dashed boxes. The present locator of edges and shades, represented by a solid box, forms a stereo pair of monocular line drawings. The dashed box marked "binocular" represents computation now operating separately to determine that pairs of points in the left and right views are "homologous", that is, representative of the same point in three-dimensional space. Binocular, or range-finding, computation will be merged with the locator of edges and shades.

At first, we called a vehicle designed to carry this system "rover". As we came to conceive of it with other senses, besides vision, and other effectors, beside wheels, we renamed it "robot".

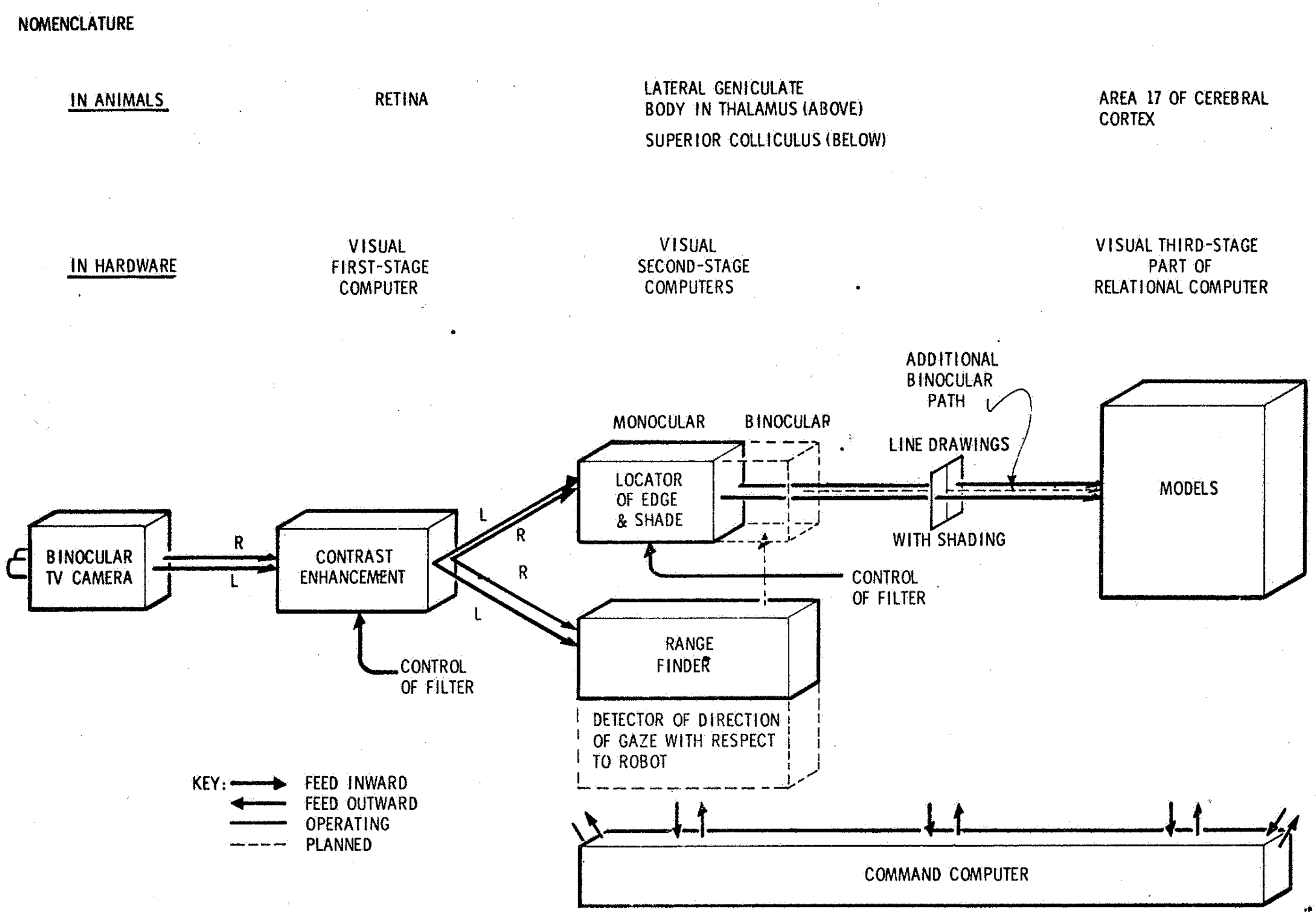


Fig. 1 Computers being developed. Feed outward for perception is indicated in the control of filters

3. BIOLOGICAL COMPUTERS

From his life-long study of the human nervous system, [5] Dr. Warren McCulloch has concluded that the essential features of its computations provide a good basis for the design of a robot. Although as a neurologist, psychologist and physiologist, he is aware of the difficulties involved in embodying mental functions in physical devices, he has nevertheless developed a simplified model of a vertebrate brain. His intention is merely to suggest an organizational structure necessary for efficient robot performance.

Figure 2 outlines his model of the vertebrate nervous system, identifying what he feels are five principal computational domains and their chief functional connections. At the left is the retina, consisting of three layers of cells, two of which seem to perform most of the computation. The eye is shown as representative of the senses because its computational capacity qualifies it as a principal computer; it is the foremost data source to the primate brain, providing two million of its three million inputs. Other senses shown are acoustic (represented by the cochlea), vestibular and somatic.

At the upper left is the cerebrum, which Dr. McCulloch calls the "great computer" and in which computation is carried out in many layers. Each of these, if unfolded from our brains, would be about the size of a large newspaper.

The computer which controls all others is shown at the center right. It is the reticular core of the central nervous system which extends from the base of the cerebrum through the spinal cord. It directs the main focus of attention so as to determine what type of activity is to occur from moment to moment. By committing the animal to one or another mode of behaviour, it controls all other computers and, through them, the whole organism.

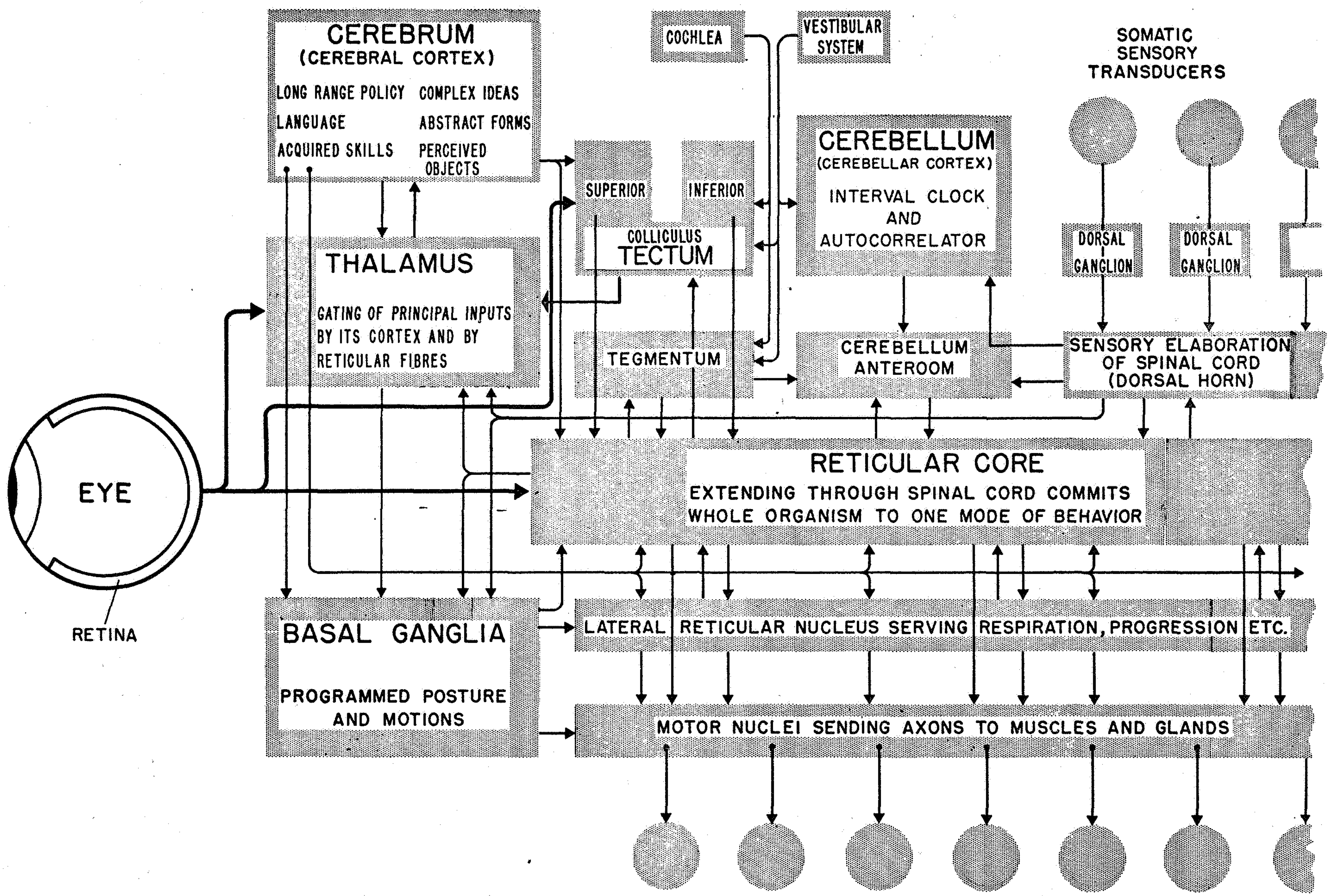


Fig.2 Block diagram of generalized vertebrate nervous system.
Feed outward paths are not shown.

Clusters of nerve cells at the base of the cerebrum comprise the basal ganglia, a computer shown at the lower left of the figure. Here are programmed all of the innate or learned total action patterns of the body, such as feeding, walking or throwing a ball. Additional programs are acquired through the growth of connections to the motor-control nerve cells, shown along the bottom of the illustration.

Completing the list of principal computational areas is the cerebellum, shown at the top of Fig. 2. It computes the termination of a movement, such as reaching to touch an object, and requires inputs from the vestibular system, to detect tilt and acceleration of the head, and from skin- and muscle-sense cells to detect posture and the nature and position of what is being touched.

Interconnected with the principal computers are switching structures, such as the thalamus, colliculus, and cerebellar anteroom. In fish, amphibians, and birds, the superior colliculus perceives form and movement; in visual mammals, it determines the direction of gaze and reports by thalamic relay the cues of seen motion to the secondary visual cortex. The inferior colliculus is concerned with auditory and vestibular inputs as well as with orientation of the body image in space. Below the colliculus is the tegmentum, which is concerned with the relations between things seen, heard, and felt and the control of progression and postural righting actions.[6]

Around the reticular core are specialized structures that could also be called computers, such as the nucleus of nerve cells that control respiration and other routine bodily functions, and the dorsal horn of the spinal cord, through which pass inputs from sensory cells. Note that the reticular core acts on all other computers and that they report to it. It reaches decisions with the aid of raw data from the sensory systems but its main input comes from the other computers.

The computers of Fig. 2 are shown as they are arranged in animals with horizontal spines. Monkeys and man have the same computers in approximately the same relation, but the arrangement is vertically distorted, with the cerebrum, now very much larger, at the top.

All these computers have a common ancestry. All evolved from the central computer, the reticular core, and in so doing have established only those interconnections necessary for efficient communication with it. Out of the reticular core has thus evolved the complexity necessary to meet the demands of the entire system.

4. AN ENGINEERING ANALOG

Figure 3 is a diagram analogous to Fig. 2, labelled with engineering terms to suggest how the animal system can be simulated. For example, in place of the retinas are the cameras and the visual first-stage computer, previously shown in Fig. 1. First stage computers receive inputs from all of the senses--auditory, vestibular and somatic sensory. Each is called a computer rather than a precomputer or preprocessor to indicate that it receives feed-outward signals from the central computers.

Other substitutions are as follows:

command computer	for	reticular core
relational computer	for	cerebral cortex
timing, coordinating and autocorrelating computer	for	cerebellum
computer of effector sequences	for	basal ganglia (nucleii)
executive computer	for	lateral reticular nucleii

The connections to the command computer shown in Fig. 3 are only those referred to in this paper. Eventually this computer will connect to every subsystem and every subsystem will connect to it. Examples of sensory subsystems are visual, auditory, vestibular, contact and kinesthetic. Examples of effector subsystems are vehicle, arms, camera focus and camera gimbals.

When the feed-outward paths are added, and control loops are drawn through the environment in the manner spoken of in Section 2, the system is seen to be composed entirely of closed loops.

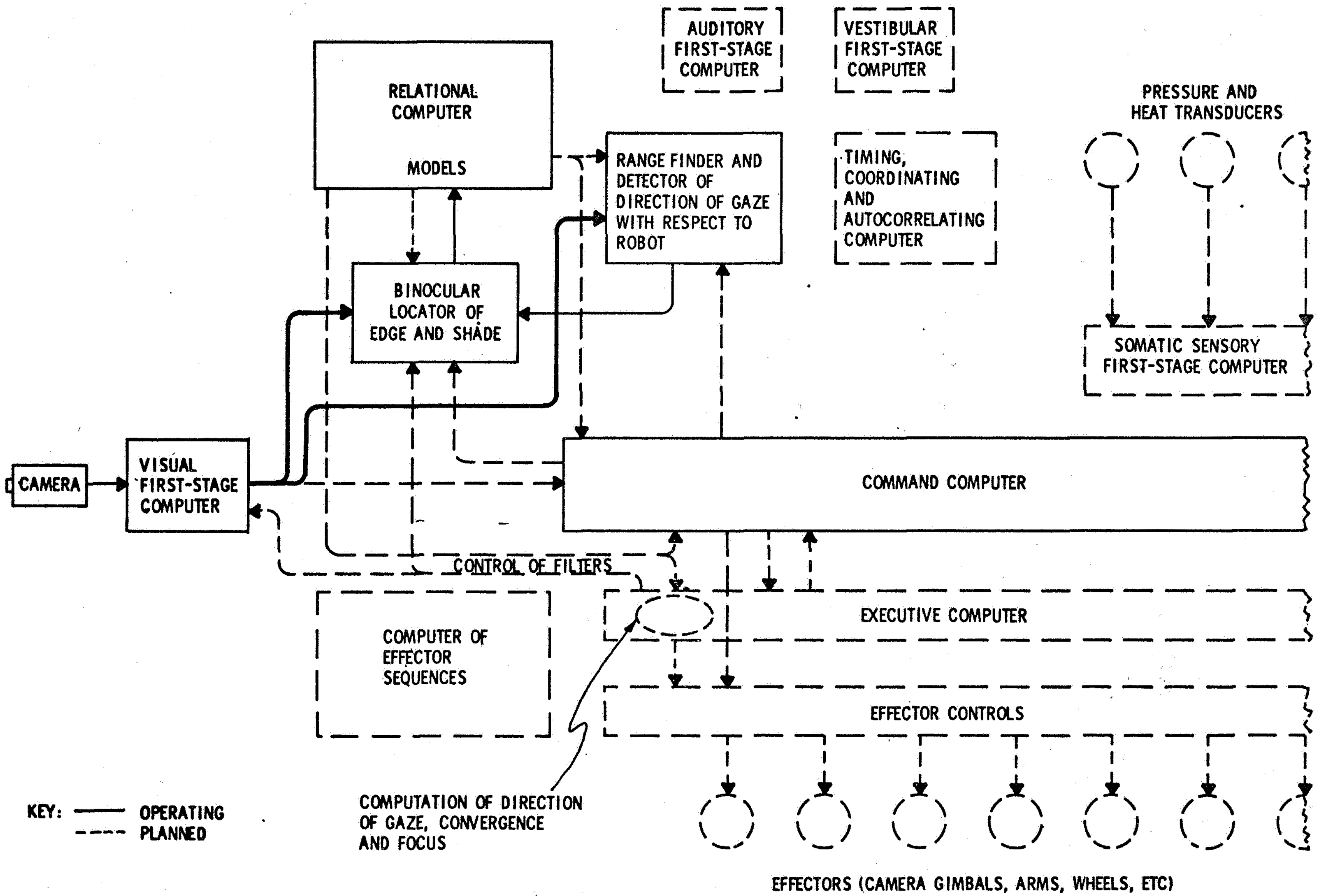


Fig. 3 Engineering analog of generalized vertebrate nervous system.

5. LOGIC IN BIOLOGICAL AND ELECTRONIC COMPUTERS

On the one hand, we have the nets of the nervous system; on the other, the contrived logic of electronic computers. In a living nerve net, branches interconnect; information from every source mixes at many places with information from every other source, and affects every output. This is called an "anastomotic net".

The electronic computers we are designing at present are not programmable general purpose (GP) machines. A GP computer is primarily intended for sequential computation on stored data. It is adept at taking data from one part of memory, modifying it, then putting it back into memory. The need here, however, is to compute on a large-volume stream of data entering from the outside. Accordingly, special purpose (SP) computers are being designed in which computation is performed on the data soon after it enters the system from one or many sources. There is no more than buffer storage between the entrance and the computation.

For each of the "five principal computational domains" described in Section 3, we aim to build an electronic approximation to an anastomotic net. To do this we need to:

1. Approximate its functions;
2. simulate these approximate functions in a configuration that is realizable in hardware and
3. realize these approximate functions in an SP computer.

For example, to design a model of a retina we first approximated its function of enhancing contrast by the function described in Section 7. We call this function a "visual first stage computer", are simulating it in a GP computer, and have partially constructed an SP computer to do it. Other functions of the retina will also be simulated such as enhancement of color contrast.

To design a command computer we first approximated the function of the reticular formation in animals, described in Section 13. We call the successive simulations S-RETIC and STC-RETIC. Design of an SP computer to perform these functions is under way.

6. MEMORY

What characterizes the evolution of S-RETIC into STC-RETIC is the addition of delays which provide the storage that is basic to memory. The same will happen in the evolution of the visual computers. They now include no memory, except for shift registers required for computations, because, for early Mars landings, the intent is only to reduce pictorial data on Mars for reconstruction and viewing on earth. STC-RETIC, on the other hand, can be conditioned to respond to a pattern of stimuli, can drop out this conditioning in the process called habituation, yet can pick it up again. These traces are stored in tables in STC-RETIC, but will be stored in adaptive elements in the hardware design. What are stored are the modes of behavior that are responses to general stimulus patterns.

Visual memory will store the relations found in these stimulus patterns by the visual subsystem. Hence the term: relational computer. An object will be stored, not as a picture, but as a structure of relations, or model, which cause the robot to do something: run from the object, pick it up, experiment with it. Such a model can either be built in or learned.

If we construct a robot, it will be primarily to perform a useful task, and only secondarily to show what is in its head. Aptitude for drawing pictures can be built in and proficiency learned. Since the camera-computer chain proposed for an early Mars landing is only part of a robot, its only useful output will be stereo pairs of line drawings with shading. If it evolves into the visual subsystem of a robot, the relations it now forms into line drawings can be in a three-space memory where they can be rotated so that they can be viewed from any direction.

We would call this memory an "associative" computer were it not that this term has a different meaning in engineering than in physiology. In engineering, it means, "content addressable", which is not an adequate memory from our point of view. As Dr. McCulloch puts it, "The memory we need should be addressed on the basis of relations, appropriate to its mode of behavior. We know _a priori_ that spatial relations, constituting objects, form categories both to guide locomotion, etc., and to form the bases of descriptions. Size and precise shape are secondary. Just as a baby has a built-in mechanism to find and follow faces, and only later to recognize particular ones, so our robot should see abstractions first and qualify them later in terms of corners, angles, surfaces and edges, as we do a face, in terms of eyes, nose, mouth and eventually ears.

"Since a relation can be described in a sentence, a computer, designed with relational addressing for visual relations, can be extended to verbal ones."

7. FIRST STAGE OF VISUAL COMPUTATION

The scene before an animal eye or a television camera can be described as a mosaic of luminances. If you doubt this, take a luminance meter, such as a photographic exposure meter, and aim it in a sequence of directions from left to right along a horizontal line; then in the same sequence of directions along a second horizontal line, below the first; then along a third and a fourth horizontal line; and so on until you have scanned a square pattern or raster. The readings of the meter are the mosaic of luminances. Inverted and exchanged left for right, this same mosaic is the image at the back of your eye or on the face of a television camera tube.

As you observe point to point across the image, you can detect change in luminance and represent each change by a dot. Sufficient dots form a line and sufficient lines a line drawing. Addition of low resolution (low spatial frequency) changes in luminance gives the drawing shading.

Whether we take animal vision as our model, as we are doing here, or develop designs independent of the animal, as others do, we find that three stages of computation are needed to achieve the abstraction which we call a "line drawing with shading" and make it useful in the command and control of a robot. As shown in Fig. 4, the first stage enhances contrast. The second stage forms line drawings which are either mapped in the third stage or, as proposed for an early Mars landing, transmitted to earth. A part of the second stage not yet tied into the sequence of Fig. 4, determines the range of dots mapped in the third stage. Still another part, to determine shading in the line drawing, has been simulated by an artist and will be automated and tied in later.

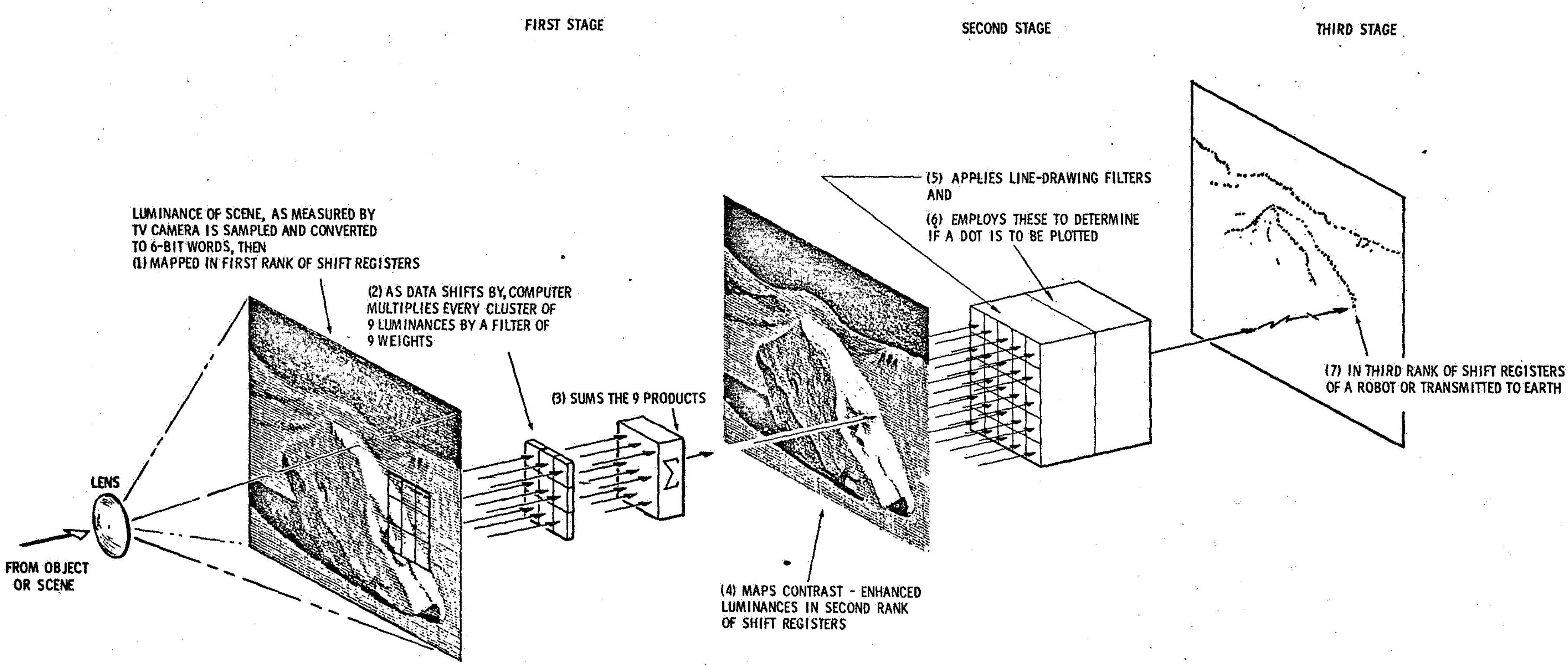


Fig. 4 Levels of visual computation performed on a mosaic of luminances. At levels 1 and 4, horizontal lines in the image represent bands of luminances; thus, the squares drawn on the image are oversize. The images are neither inverted nor turned right for left as they should be.

The stages presently operating as a sequence are broken down into levels in Fig. 4. Continuous luminance measurements made by the TV camera are sampled, converted to 6-bit digital words, and, in level 1, mapped. At level 2, parallel computation is performed on a number of luminance measurements which, for illustration purposes, is shown as 3 x 3, although in present experiments, it is 9 x 9.

The Jet Propulsion Laboratory (JPL) of the California Institute of Technology has improved, by computer, the quality of pictures sent back from the moon and Mars and x-ray radiographs of medical cases. Their objective is "to make selected features easier to see. This might require suppression of useless data such as random noise and background shading or perhaps amplification of fine detail."[7]

Our first objective on the other hand, is to reduce pictorial data for both transmission from Mars to earth and for reconstruction there. Only after it has gone through reduction and transmission do we want to make it easier to see. Our second objective is to reduce pictorial data to enable a robot to see. Yet our objectives and JPL's appear achievable in the same way, namely, by operations on the spatial frequencies in the image.

The output of a TV camera is a waveform and as such is analyzable into frequencies of luminance amplitude in the horizontal direction of sweep of the camera beam. Since the TV raster is made of many lines, measured vertically, the image on the face of the camera tube is also analyzable in the vertical direction. The frequencies of luminance amplitude in all possible directions within the plane of the image are called "spatial frequencies."

Our equipment operates on these frequencies and amplitudes by employing digital filters, although the filters can be "analog" in the sense that computer designers use this term. A digital filter is a matrix of weights which can be made to operate on a matrix of luminances (the image) in the following way: Separate the matrix of luminances into small overlapping arrays called sub-matrices, convolve each sub-matrix with the filter and then sum. We will explain "convolve" as we do it.

To amplify fine detail, a filter should pass high spatial frequencies, reject low spatial frequencies and noise, and do this in all possible directions in the plane of the image. Eq. 1 shows a simple filter designed for this purpose, convolved with a sub-matrix of uniform luminances.

$$\underset{\text{Luminances}}{\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}} \star \underset{\text{Filter } W_1}{\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}} = 0 \qquad (1)$$

To convolve, as the star requires, multiply each number in the filter by its corresponding number in the sub-matrix of luminances and sum the product. Since this convolution sum is zero, adding it to the central luminance in the sub-matrix of luminances produces no effect. That is, there is no fine detail. Given a sub-matrix of different luminances and the same filter, the convolution produces:

$$\underset{\text{Luminances}}{\begin{bmatrix} 1 & 2 & 2 \\ 1 & 2 & 2 \\ 1 & 2 & 2 \end{bmatrix}} \star \underset{\text{Filter } W_1}{\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}} = 3 \qquad (2)$$

Adding this 3 to the central luminance, in this case, 2, would enhance the contrast between the central luminance and the 1 at its left.

---

*"Luminance" from now on is used to represent luminance measurement.

By extending the band of 1's in the above sub-matrix of luminances, indefinitely up and down and to the left, and the band of 2's indefinitely up and down and to the right, we obtain a full matrix of luminances. The dots represent numbers continuing outward:

$$\begin{bmatrix} & & \cdot & \cdot & \cdot & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \cdot & \cdot & & \\ \cdot & \cdot & 1 & 1 & 2 & 2 & 2 & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 & 2 & 2 & 2 & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 & 2 & 2 & 2 & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 & 2 & 2 & 2 & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 & 2 & 2 & 2 & \cdot & \cdot \\ & & \cdot & \cdot & \cdot & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \cdot & \cdot & & \end{bmatrix}$$

If we convolve the filter W_1 with all of the possible overlapping 3 x 3 sub-matrices in this matrix and then sum, we obtain the following matrix:

$$\begin{bmatrix} & & \cdot & \cdot & \cdot & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \cdot & \cdot & & \\ \cdot & \cdot & 0 & -3 & 3 & 0 & 0 & \cdot & \cdot \\ \cdot & \cdot & 0 & -3 & 3 & 0 & 0 & \cdot & \cdot \\ \cdot & \cdot & 0 & -3 & 3 & 0 & 0 & \cdot & \cdot \\ \cdot & \cdot & 0 & -3 & 3 & 0 & 0 & \cdot & \cdot \\ \cdot & \cdot & 0 & -3 & 3 & 0 & 0 & \cdot & \cdot \\ & & \cdot & \cdot & \cdot & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \cdot & \cdot & & \end{bmatrix}$$

Adding this to the full matrix of luminances, at the top of page 19, we obtain:

$$\begin{bmatrix} & & \cdot & \cdot & \cdot & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \cdot & \cdot & & \\ \cdot & \cdot & 1 & -2 & 5 & 2 & 2 & \cdot & \cdot \\ \cdot & \cdot & 1 & -2 & 5 & 2 & 2 & \cdot & \cdot \\ \cdot & \cdot & 1 & -2 & 5 & 2 & 2 & \cdot & \cdot \\ \cdot & \cdot & 1 & -2 & 5 & 2 & 2 & \cdot & \cdot \\ \cdot & \cdot & 1 & -2 & 5 & 2 & 2 & \cdot & \cdot \\ & & \cdot & \cdot & \cdot & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \cdot & \cdot & & \end{bmatrix}$$

Thus where contrast exists, the high luminance has been made higher, the low lower. Both of the above methods of amplification or enhancement are called "lateral inhibition" because, while the center of the matrix is excitatory, the periphery is inhibitory.[8] The bands of enhanced light and dark that result from application of the filter are called Mach bands after Ernst Mach who first described them.[9] Our retinas perform this operation so that we see Mach bands wherever there is a steep step in luminance.

The multiplications in the above operations are pictured in Fig. 4 as taking place in level 2, the summation in level 3. To combine the two levels of computations we can double the central weight in W_1. Furthermore, to keep most convolution sums in scale we can multiply the filter by 1/8. The new filter can be described as such that the central weight equals twice the absolute value of the surround and the sum is one:

$$W_2 = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 16 & -1 \\ -1 & -1 & -1 \end{bmatrix} \times 1/8 = \begin{bmatrix} -1/8 & -1/8 & -1/8 \\ -1/8 & 2 & -1/8 \\ -1/8 & -1/8 & -1/8 \end{bmatrix} = 1 \qquad (3)$$

By extending the above reasoning to include consideration of noise, Jerome Lerman devised the 9 x 9 filter of Appendix A.2 which the computer then employed to produce the enhancement in contrast shown in Fig. 6a.* While this figure does not show the enhancement as clearly as the original read from the computer, the improvement can be seen along the right side of the rock and around the base of the hill at the left. Where there is contrast between adjoining luminances in the original image, enhancement makes the dark side of an edge darker and the light side lighter, thus forming Mach bands along the edges of the rock, the stick, the hills and the crater. Figure 6b shows the effect of enhancing the contrast of Fig. 6a.

A way of keeping the display on scale when applied to spots and edges of maximum contrast is given in Appendix A.1.

The vertical lines in the displays of Figs. 5 and 6 are due to the method by which the computer of Fig. 8 acquires pictorial data. As the electron beam of the camera scans each horizontal line of the raster, the computer commands the reading of one luminance measurement.[10] The computer employs the same command as long as it can with the result that successive measurements are below each other on vertical lines. The unevenness in the spacing of vertical lines is due to drift of one vertical line with respect to another in the display scope.

*A stereoscope for reviewing this illustration may be obtained from Air Photo Supply, Yonkers, N.Y. Adjust it to your interocular distance. If you do not know this distance, use 63 mm.

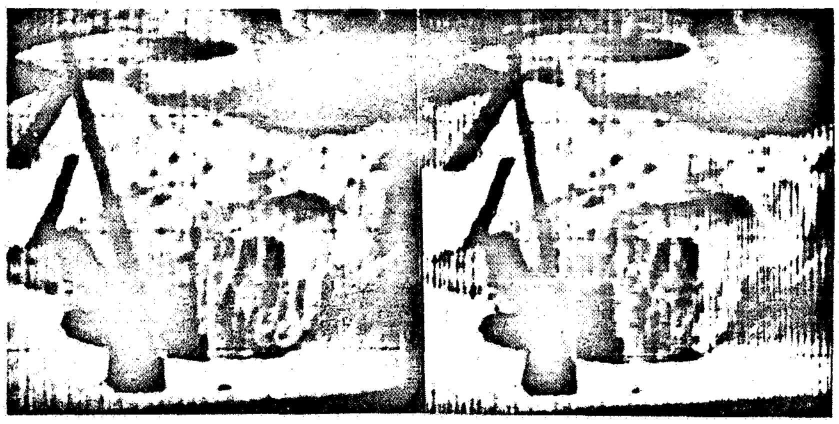


Fig. 5 Scope displays of images acquired by TV camera.

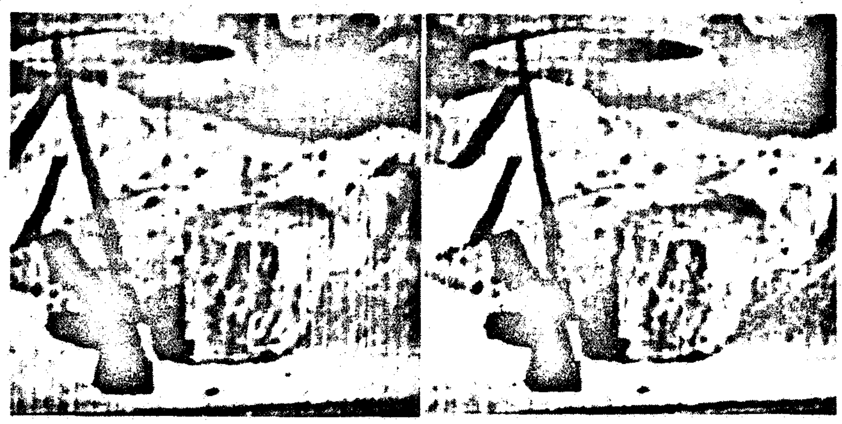


Fig. 6a Images of Fig. 5 contrast-enhanced once.

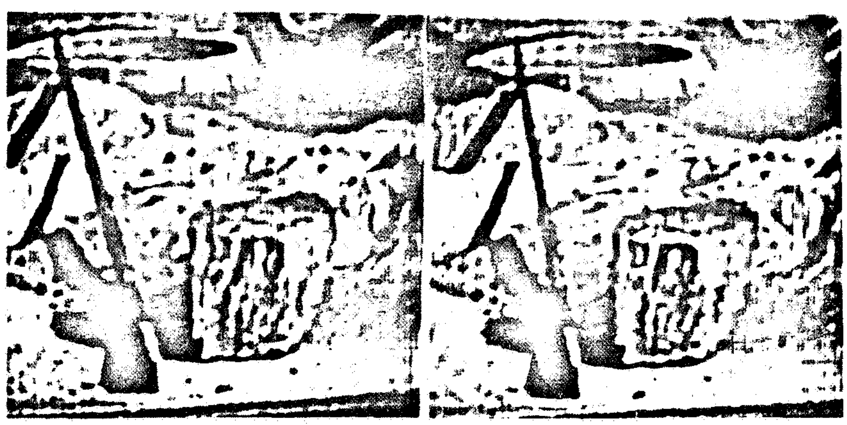


Fig. 6b Images of Fig. 5 contrast-enhanced twice.

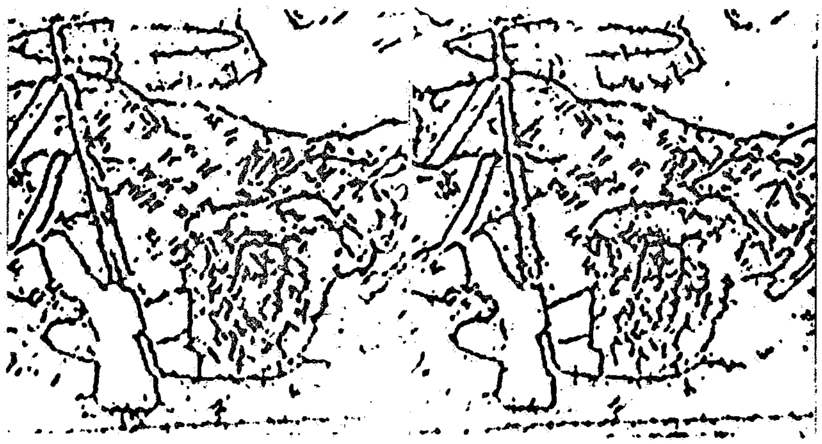


Fig. 7 Line drawings formed from images of Fig. 6a.

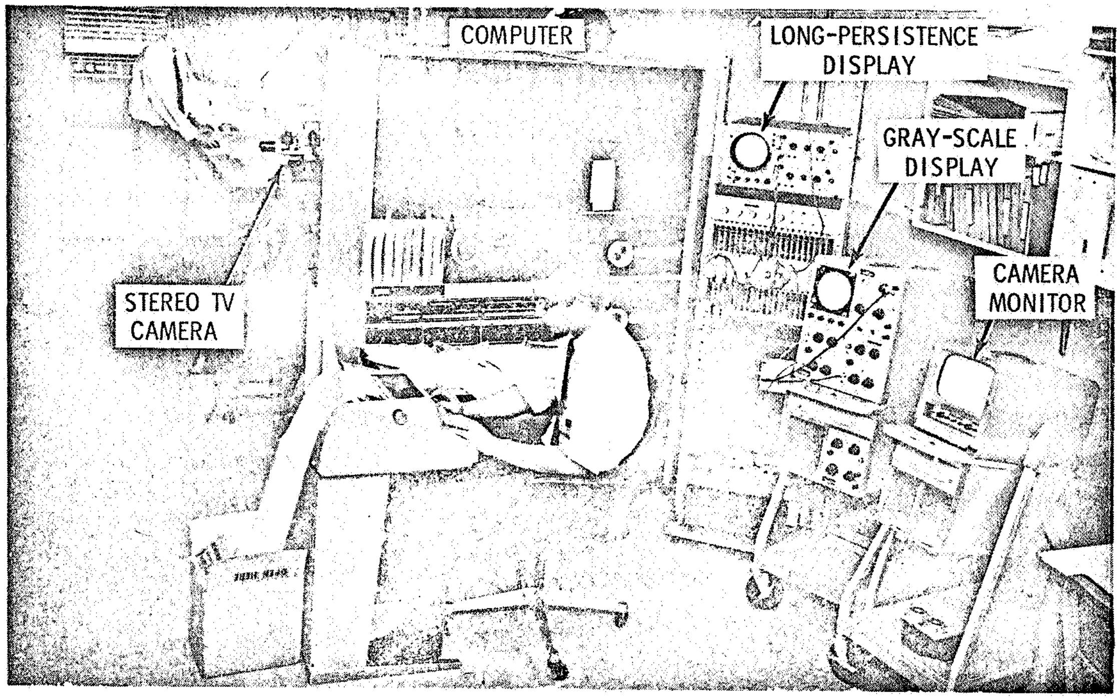
8. SECOND STAGE OF VISUAL COMPUTATION

The second stage is being designed initially to reduce the contrast-enhanced image of a scene to a line drawing with shading, and later to permit perception and recognition. For the latter two purposes, filters can be adjusted, as the image moves past them, to seek a match with features in a stored model.

Thus, feature detection can be varied, not only for the entire image, but for each matrix within the image, under control of either a remote human operator or a local command computer. This position-by-position control of the processing of the image, represented by the leftward arrows in Fig. 1, separates our work from that at JPL. It makes perception possible.

Filters, to be described in a later report, extracted from the data of Fig. 6a the line drawing of Fig. 7. This line drawing does not contain enough information either for a scientist on earth to judge what is being pictured or for adequate automatic comparison between incoming and stored data in the process of perception. However, by the addition of low-resolution luminance data to left and right views, and presentation of the two views stereoscopically, there may be enough information. Figure 10 is an example of levels 5 and 6 computation alone on data received from the scene pictured in Fig. 9. Figure 11 shows coarse-resolution measurements of luminance on a scale from 0 to 7, which an artist employed to paint in the swatches of gray in Fig. 12. When this reconstruction is performed by computer, it will illustrate how the appearance of a Martian scene can be reduced for transmission from Mars to earth and then reconstructed on earth for viewing there. The data reduction here is by a factor of 30.

The stereo pair of views shown in Fig. 5 was not taken with the mirrors illustrated in Fig. 7, but by taking one view at a time and moving the camera between takes. This is simpler for a report.



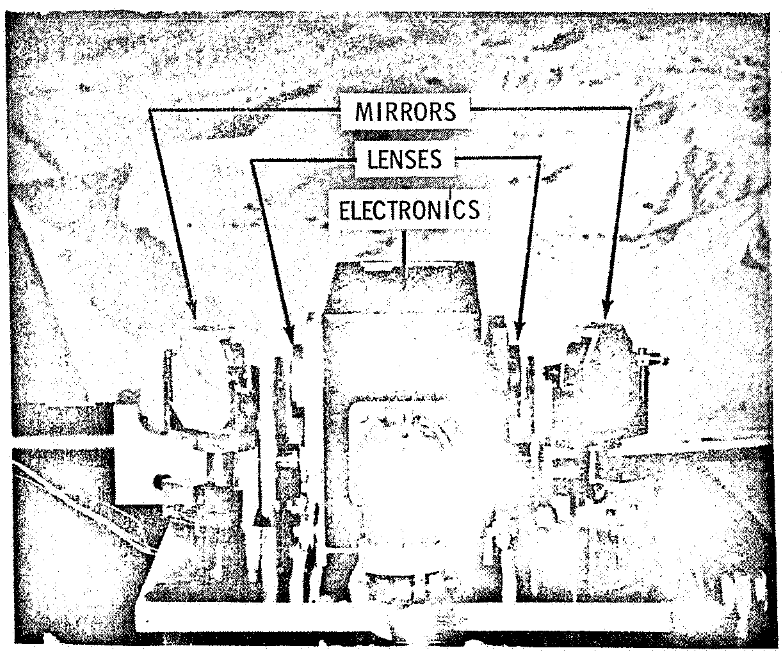


Fig. 8 Equipment for simulating light-weight, low-power hardware;
(above) camera-computer chain for simulating visual computers,
(below) binocular, or stereo, TV camera.

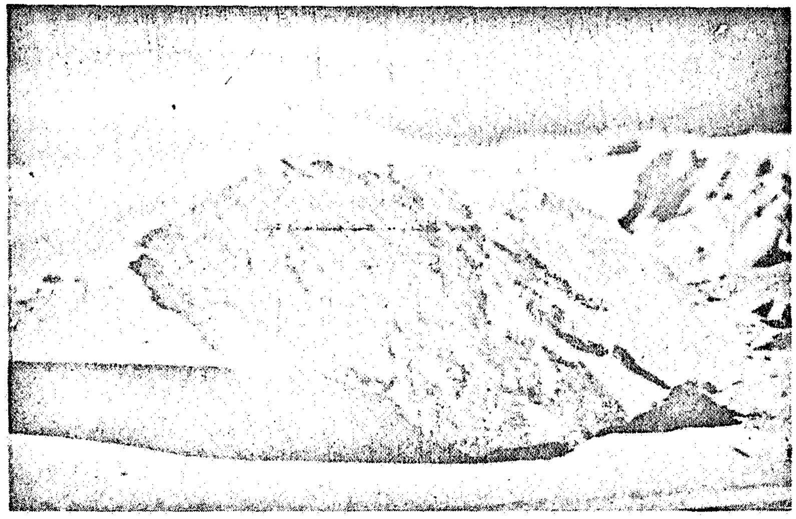


Fig. 9 Photograph of scene to be formed into line drawing with shading.

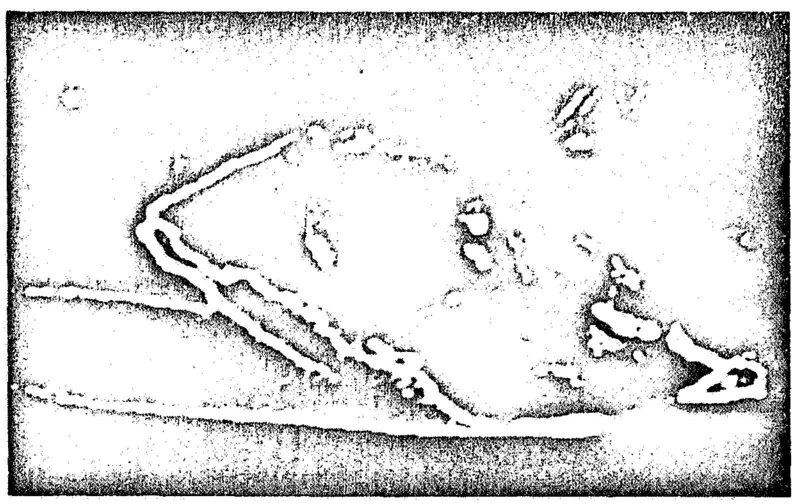


Fig. 10 Computer-generated line representation of scene.

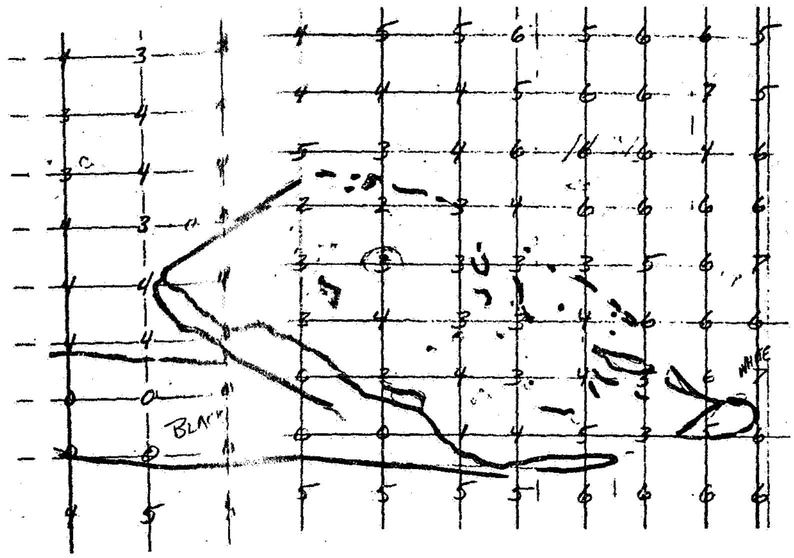


Fig. 11 Grid of scene luminance values superimposed on Figure 10.

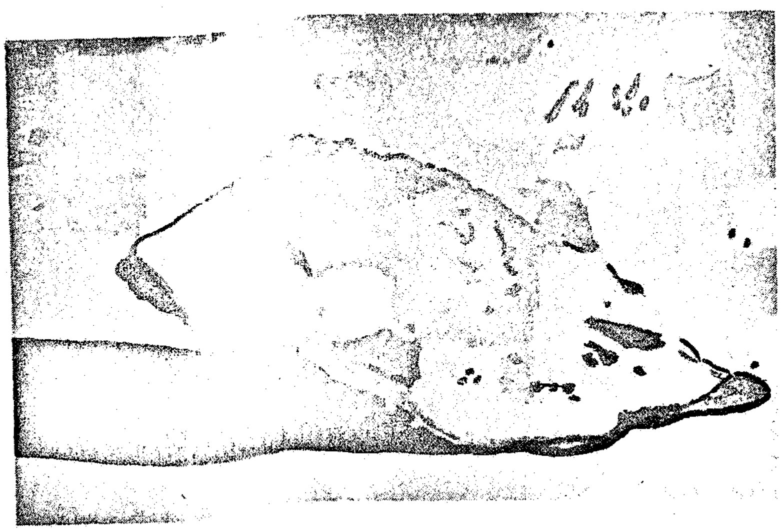


Fig. 12 Painting with shades of grey prescribed by the luminance values of Figure 11.

9. HARDWARE VERSION OF VISUAL FIRST-STAGE COMPUTER

We designed the computation first so that it could be implemented in light portable hardware: then we simulated this hardware in the computer of Fig. 8 and achieved the results described above. The hardware design, diagrammed in Fig. 13, is inspired by the layered structure of the animal retina and laterial geniculate body of the thalamus. Since it is not practical to represent in hardware the large number of cells of these animal structures, only a cluster of cells of each layer is represented and data are moved past the cluster by shift registers.

A camera containing a single vidicon, that receives left and right views from mirrors, is shown in preference to two cameras because the former arrangement leads to much less uncertainty between the two optical paths[10].

In Fig. 13, left and right images of an object such as a rock are projected by the optics onto the face of the vidicon. Converted to digital form, the signals enter shift registers (1) which move the images past computing elements (2) and (3), which represent clusters of living cells. The filter planned for levels 2 and 3 is shown schematically at the upper left of Fig. 13 and is given in detail in Appendix A.2. It consists of a central weight strongly positive, immediately peripheral less positive weights and more peripheral weakly negative weights.

The images are advanced from left to right in the shift registers at the same rate that the tip of the electron beam in the vidicon advances. As data reach the right end of the top row, they are transferred to the left end of the next lower row. In this illustration, when the electron beam of the camera tube has swept ~~13~~ 9 lines of the raster, the shift registers in the bank are full. From then on, for each new position of the electron beam, computations take place in the box behind the shift registers, and one digital word is discarded from the right end of the ~~13~~9th row of shift registers.

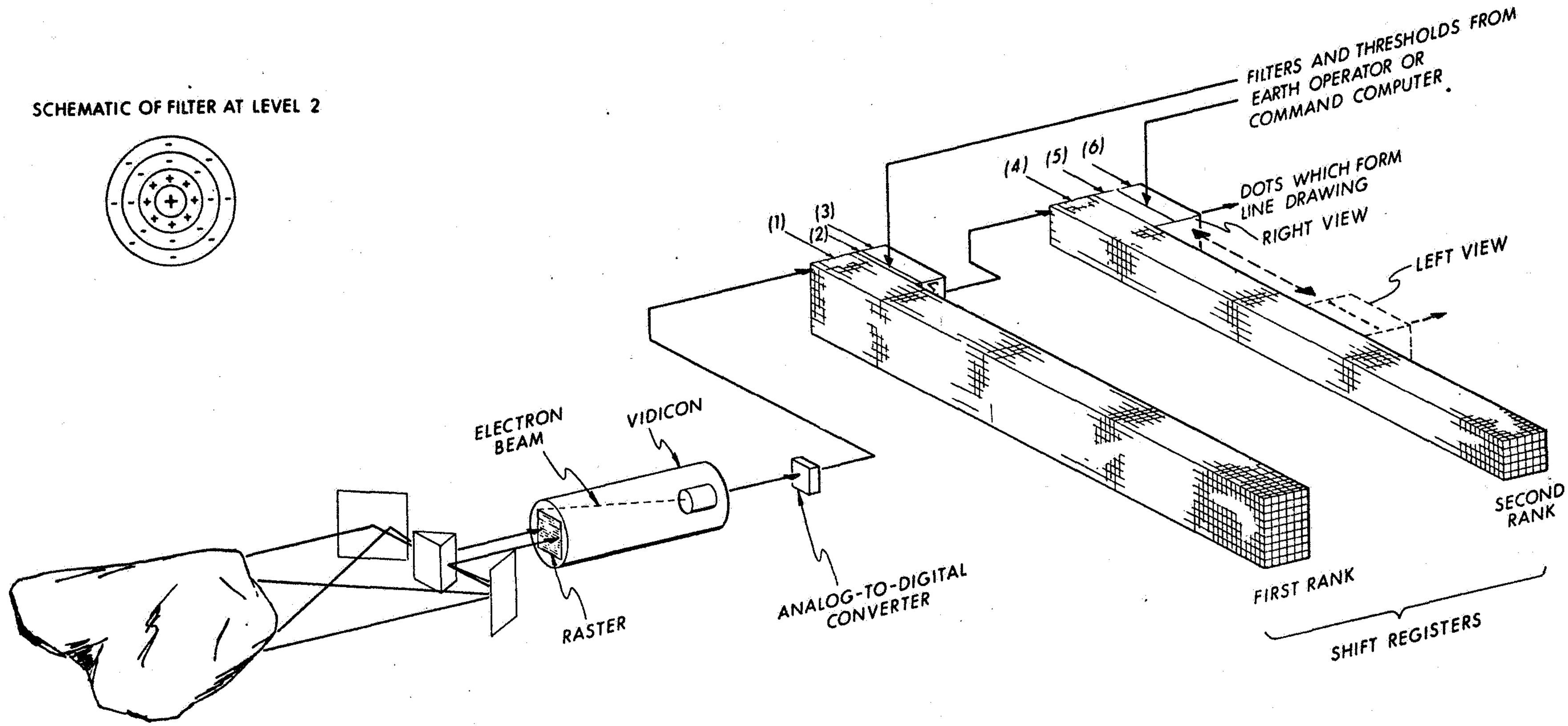


Fig. 13 Diagram of hardware to perform contrast enhancement.

Figure 14 shows test hardware under construction to perform levels 1, 2 and 3 computation, on five lines of the raster. Each of the lower five panels contains a shift register 6-bits deep. The registers shift the data past the computing element in the top panel. The medium-scale integrated circuits to be employed here, together with their wiring, can be packed into about 50 cu. in. With large-scale integrated circuits the volume could be 10 cu. in.[10]

10. COMPUTATION OF RANGE

To perform second-stage computation, either a man or a robot needs a view from a second position. We refer to the two views as "left" and "right" since they are usually taken from the same horizontal baseline. If the levels of robot computation pictured in Fig. 4 are for the left view, then either a second series of levels is needed for the right view or the levels need to be widened to accommodate both views. We have taken the latter approach in the design of the hardware.

Animals compute range from comparisons of left and right images, at several levels of computation, and from the angle between the axes of the two eyes, called the "convergence angle".[11] The equipment shown in Fig. 7 has been programmed to automatically compute range by comparing areas, in the left and right level 2 images, called "windows" (see Fig. 15). To find a pair of windows that are viewing the same object, a window is first fixed on say, the left view, according to some criterion such as the presence of an edge; then a likely area is located in the other view. Since this area contains as many potential windows as resolution elements along the horizontal axis, the problem is to determine which of these windows corresponds to the one fixed in the left view. The simplest way is to compare luminances in the fixed window with corresponding luminances in the likely area, determine the difference and use this as a criterion to decide when a best match is obtained From the data of the best match, range is computed by triangulation.

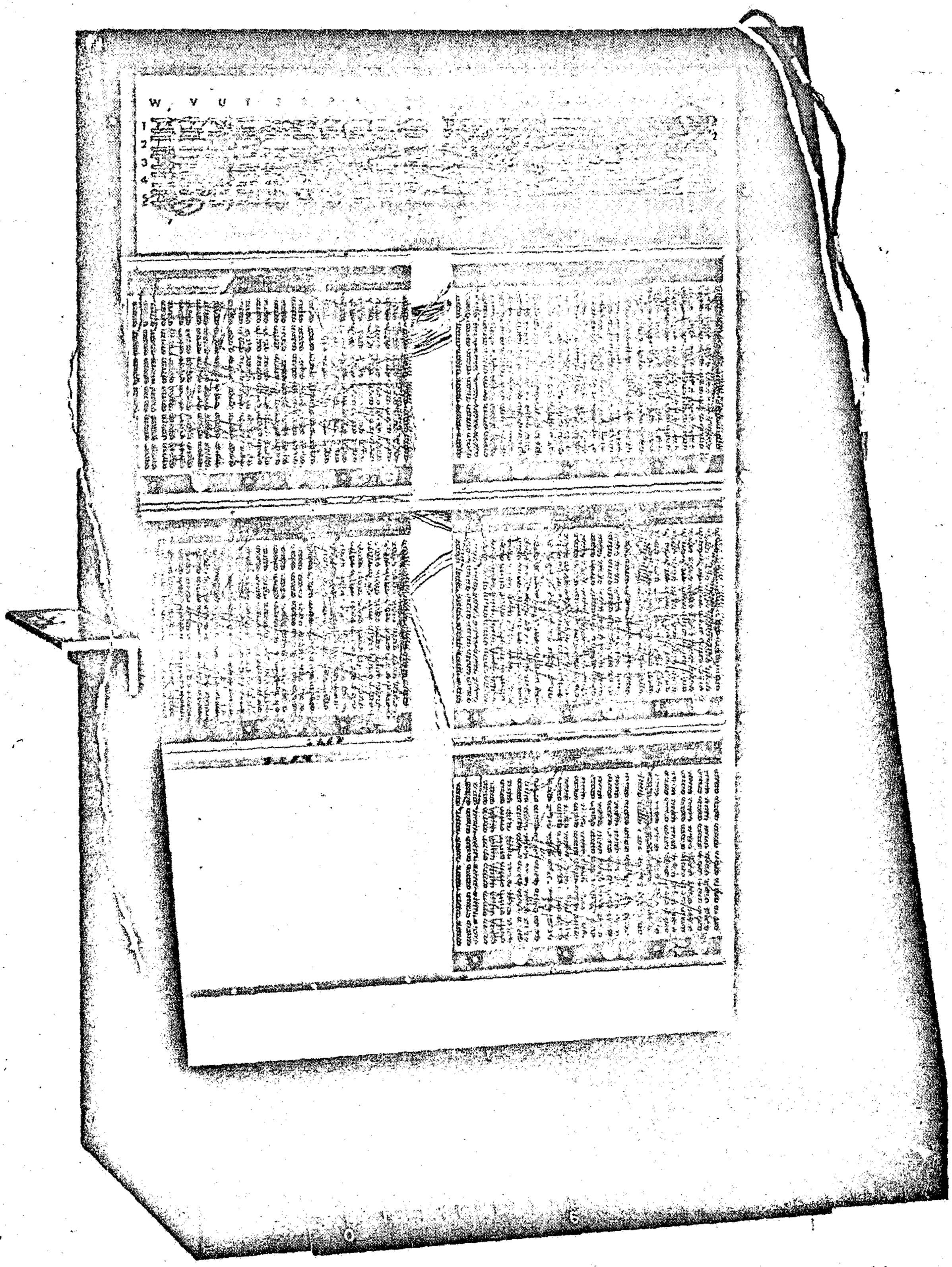


Fig. 14 Visual first-stage computer under construction. The top panel is the computing element. The other five are shift registers. Integrated circuits plug into the far side.

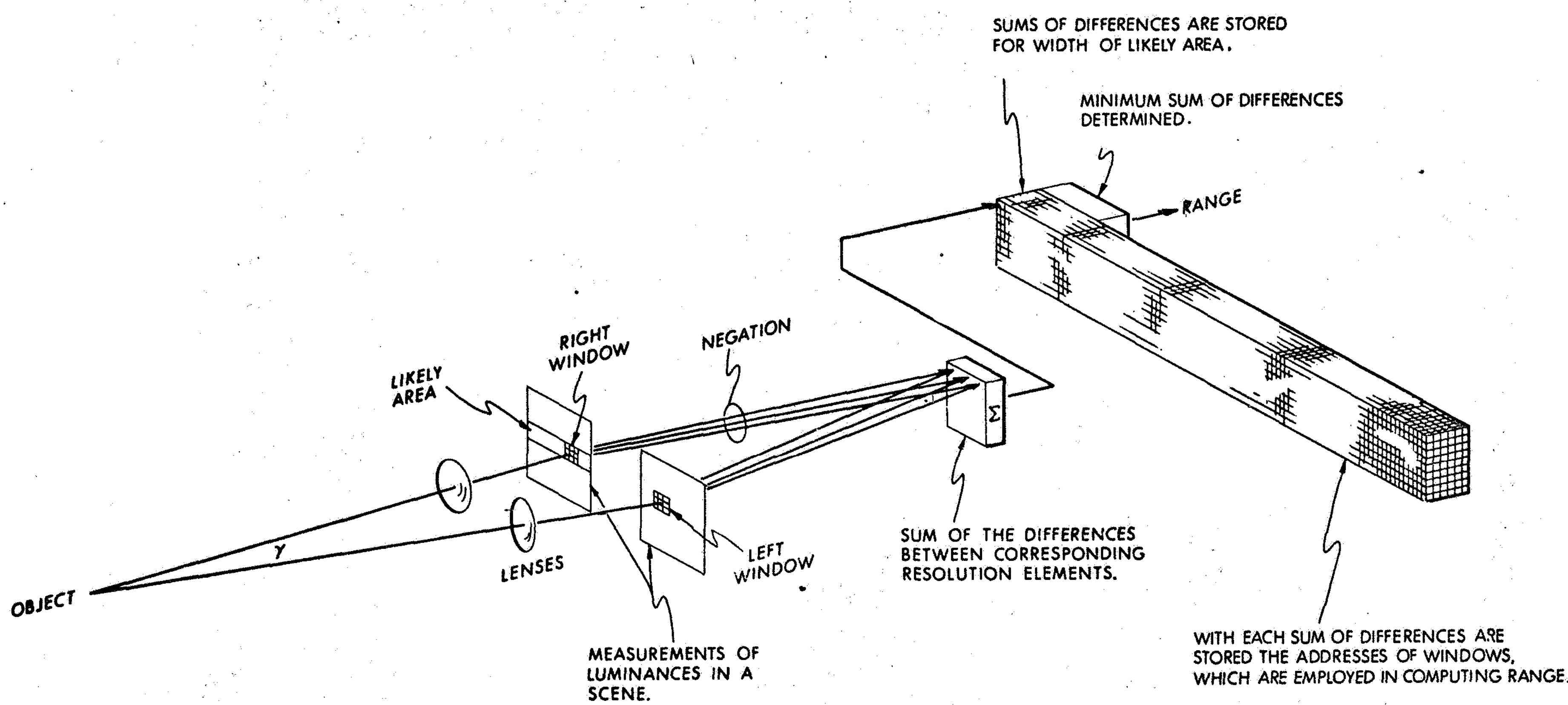


Fig. 15 Comparison of left and right views to determine range.

Employing this method, equipment shown in Fig. 7 automatically explores a likely area to determine the range of an object at 20 feet with an uncertainty of 1.5%. To perform the comparison over the entire likely area requires 16 seconds. The comparison will be performed over less area if the robot visually follows around the edge of an object or visually explores increasingly deep into the scene. In these latter cases, the visual subsystem of the robot starts with a known range and reaches out from it.

11. PERCEPTION

In the robot we plan, the command computer, assisted by the relational computer, will determine what is seen by setting filters and thresholds in all stages of visual computation so as to match an internally-generated image with an incoming one. This "Keeping up to date the internal organizing system, that represents the external world, by an internal matching response to the current impact of the world upon the receptive system" is called "perception."[12] "In a sense, the internal organizing system is continually making fast-time predictions on what is going on out there and what is likely to happen out there, and it takes anticipatory action to counter the small errors that might threaten its overall stability."

A line drawing with shading, transmitted to earth for a scientist to view, aids his perceptual process, giving him clues about the presence of objects about which he can then demand more information. Within a robot, however, a line drawing with shading will be the result of interaction between the relational computer, setting filters and thresholds, and the second stage visual computer where these filters and thresholds are tried on incoming data. When equipped to perceive, a robot will make fast-time predictions, possibly as detailed as the computer-generated image of Fig. 16 Our general purpose computer formed Fig. 16 from the equation of a cylinder, its diameter, height and illumination. It appears that perception of the cylinder could take place in the first and second stages of visual computation if the filters there are continually changed, as data are shifted past them, to search for predicted luminances in each part of an internally-generated image.

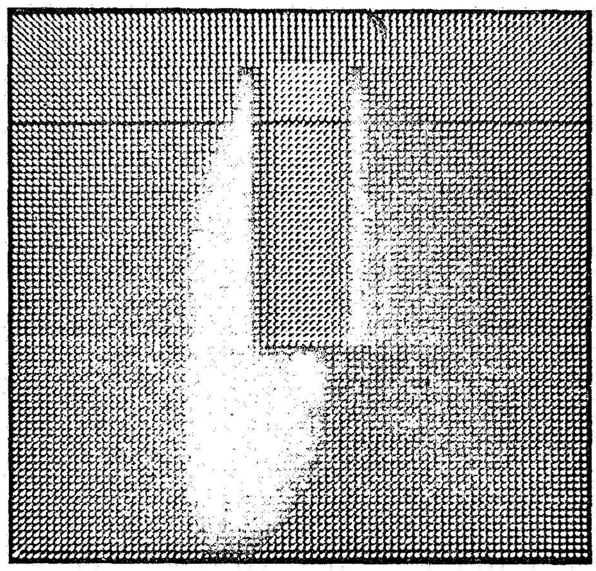


Fig.16 Picture generated by computer preparatory to experiments in perception.

12. THE CONCEPT OF A COMMAND COMPUTER

The purpose of a command computer, in an animal or robot, is to commit the entire system to one of a dozen or so mutually exclusive modes of behavior. An animal requires such a computer because it cannot "fight, go to sleep, run away, and make love all at once." (14)

Our study of a possible Mars robot indicates that it, too, can only enjoy one mode of behavior at a time. Possible modes of such a robot are:

1. Look and feel
2. Advance
3. Retreat
4. Right itself if overturned
5. Maintain itself
6. Chart its environment
7. Rest

Perform incompatible experiments as follows:

8. Experiment A
9. Experiment B
10. Experiment C

"Look and feel" is a separate mode from "advance" because the robot must be stationary while either its camera or its arm is employed.

By "chart its environment" we mean that the robot, after advancing (or retreating) an appropriate distance, will establish what a surveyor calls a "station", and mark it with a transponder. Having determined the distance from the previous station and measured the angle between this and the second previous station, the robot can form a triangle in its memory to add to other triangles previously formed. By building triangle upon triangle, the robot establishes a grid with which it can determine how far it has gone and in what direction. Through this grid it can later pass, to return to points it has already visited.

Within each mode are the detailed sequences we call acts. Advance, for example, can be either slow, fast, or by turning. These details can be developed after the command computer has been designed to choose among the above modes.

The command computer should be capable not only of selecting the mode of behavior, in response to inputs from the environment, and the other computers, but of changing the way it responds to those inputs. Ability to change the record of conditions under which it selects a mode is called "plasticity" and is exemplified by conditioning and habituation.

The first simulation of a command computer represented only its mode-selecting ability. It was called S-RETIC where S stood for its ability to respond to both its present internal state and its spatially structured input.[15] The second simulation is called STC-RETIC where T stands for its ability to be influenced by temporal aspects of the environment and C for conditioning and habituation. The properties of STC-RETIC together with several new features are now being desighed into special hardware to be called H-RETIC.

The inputs to each of these RETICs represent connections from such subsystems as the visual, described in part above, and the contact, described in part in Section 16. The number of input channels to the present RETICs is very much smaller than will be required eventually from, say, the visual subsystem of the robot. In fact, the number of input channels, (γ_1, to γ_{42} in Fig. 17) is only as many as needed to demonstrate the principles of operation.

13. HOW AN ANIMAL IS COMMANDED

In the core of the reticular formation of animals (retic), the selection of a mode of behavior is made by a matrix of fan-shaped cells embedded in regions that are stacked like poker chips. The input processes of each cell, its dendrites, give it both its fan-shape and its poker-chip-like thickness. Its output is an axon which traverses the long axis of the chip-like stack. Each cell in general receives signals from several parts of the brain, from many other reticular cells, especially those in its own neighborhood, and from several interoceptive and exteroceptive sensory pathways. Collectively, these cells decide which mode of behavior the animal will enter. In its sharpest form, this assertion is only a hypothesis; but broadly speaking it is an evident biological fact.

The informational organization of retic is analagous to a board of medical doctors who must decide upon the treatment each of a series of patients must receive. Suppose there are twelve doctors on the board, each a generalist, as well as a specialist in a different field of medicine, and that they must select one of four possible treatments. Their deliberations resemble the process by which S-RETIC selects a mode of behavior.

Like the board of medical doctors, the command computer (retic) must commit its charge to a mode of behavior which in most cases is a function of the information that has played upon it only over the last second or so (signals indicating mission are part of this). It receives information that is vast in amount, but highly correlated between input channels and arrives at one of a small number of mutually exclusive modes of behavior in a dozen or so time steps, with minimal equipment and maximal reliability. After a mode is decided upon, it must send out control directives to the other agencies in its charge to turn them to their respective tasks. Finally, that part of the command computer which at any given time has the most crucial information has authority over the mode of operation.

14. FIRST SIMULATION OF A COMMAND COMPUTER, S-RETIC

Like retic, S-RETIC resembles a stack of poker chips, but each chip is now a computer module, M_i, and represents many retic cells. (See Fig. 17) Together the modules of S-RETIC decide on a mode of behavior in response to data from an overall environment that is static while each decision is being made. Note the word "overall". A major part of the environment of retic is the cerebral cortex where are stored the plans and goals of the animal. Although S-RETIC has only 42 binary input lines and chooses among only four modes of behavior, it demonstrates principles that can be applied on a much larger scale for a robot.

The 42 input lines, λ_i, are connected from sensory subsystems, S_i, to modules, M_i, in several-to-several, but not all-to-all fashion. The outer box in Fig. 17 represents a generator to simulate an environment formed in response to the input, Σ. At

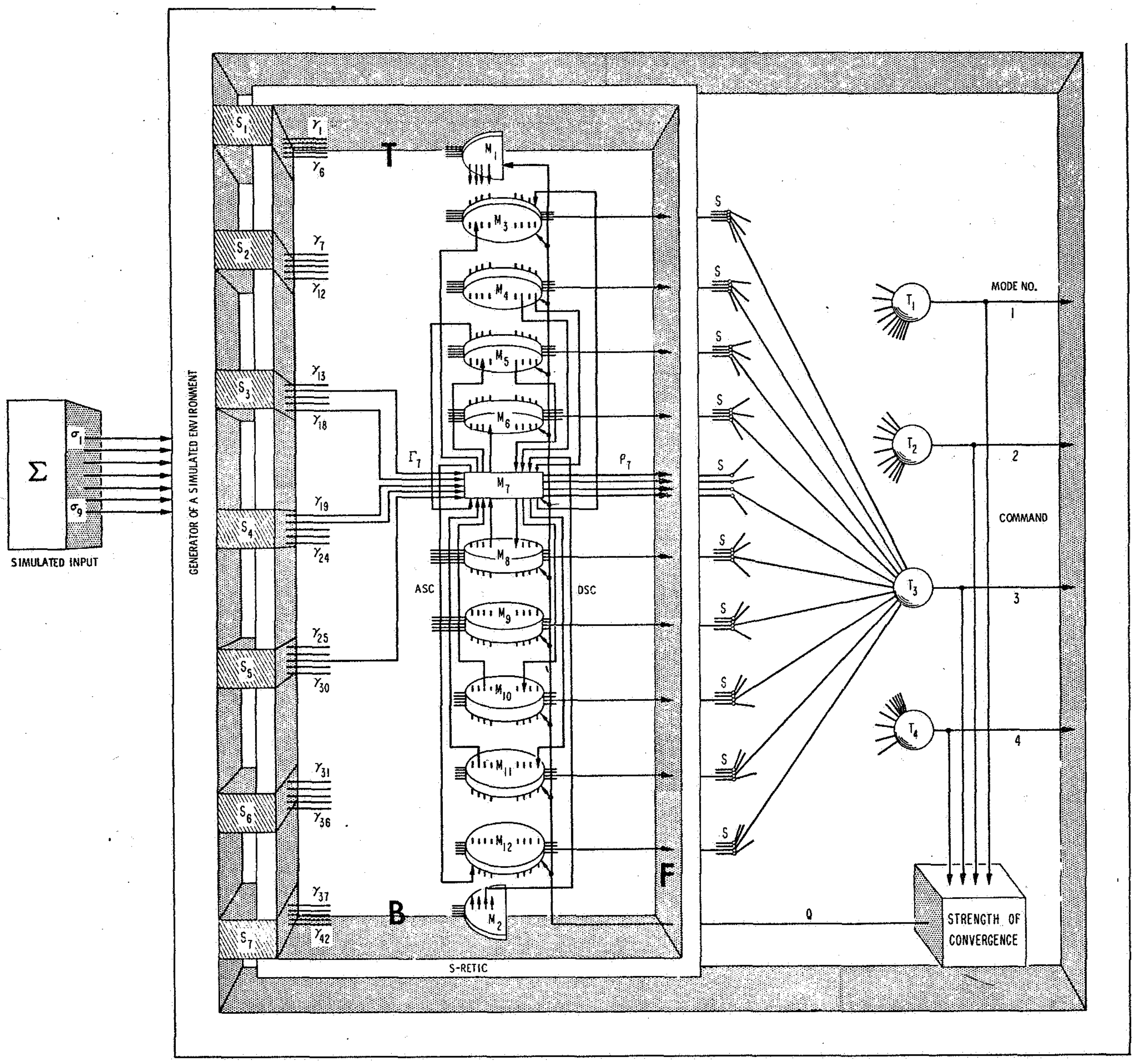


Fig. 17 Simulation model of the command computer, S-RETIC, and threshold units (T_i) that determine convergence. The M_i are logic modules; the S_i are sensory systems; all P_i (only P_7 shown) are modular mode-indicating output lines; Σ simulates a RETIC environment that engenders input signals (sights, sounds, etc.); T and B are the top and bottom boundaries of S-RETIC; asc and dsc are the ascending and descending "nerve" bundles. For clarity, the connections that recur on all M modules are shown only on M_7.

this stage of design, all σ_i and γ_i lines carry binary signals. All of the other lines into and out of modules carry numerical estimates of probabilities.

Each of the 12 logic modules computes from its input information the probability that each mode of behavior is best for the robot. After this initial guess, the modules communicate their decisions to each other over the low capacity ascending and descending lines. Then each module combines this information in a nonlinear fashion with new information coming into it to arrive at adjusted probabilities that each of the four modes is best for the robot. The module, in turn, communicates this adjusted guess to the modules to which it is connected above and below. The process repeats until six or more modules find one mode of behavior best for the robot with a probability greater than 0.5. This threshold is sensed by threshold units T in remote motor controls.

Each try at consensus, or convergence, is called a cycle of operation. For details see Appendix B.

S-RETIC is a computer program operated now as part of the larger program described below. It always converges to the desired mode of behavior in less than 25 cycles, but 30 are allowed for it in a larger time period called an "epoch" with the new model described below.

15. SECOND SIMULATION OF A COMMAND COMPUTER, STC-RETIC

In the second simulation of a command computer, the already-operating S-RETIC was expanded to provide interconnections between the a-parts of the modules (ω lines), short- and long-term memories (STM and LTM) in each a-part and channels through which the experimenter can reinforce each module. In addition, the number of γ lines to each module was increased to seven; the σ lines were increased to 13, and the mode of behavior of the robot (RM) was fed back to each a-part.

The new model is called STC-RETIC where S and T stand for a spatially and temporally changing environment, C for conditioning and habituation. Where each module of S-RETIC has a transformation network to form its initial guess, each module of STC-RETIC draws its initial guess from its LTM. During this and the ensuing epoch, it computes the effects of reinforcements, given it by the operator, and then adjusts its LTM accordingly. The result is conditioning, habituation and other forms of "plastic behavior". For details see Appendix C. Given there are examples of the Pavlovian conditioning of a robot in a remote environment and the dropping out of this conditioning, called habituation. Development is also presented there as a form of plastic behavior.

The next step is to design a computer to be both a refinement of STC-RETIC and a more faithful reflection of the neurology of the core of the reticular formation. H-RETIC as it will be called, where H stands for hardware, will be organized much like the STC-RETIC with physically separate modules containing adaptive elements for memory.

16. CONTACT SUBSYSTEM

Design of a sense of touch has progressed to the point of planning a hand and arm to reach out and press lightly against surfaces to the front and side of the robot. Figure 15 shows the tactile hand with a single finger. A grasping hand is also being designed to be carried by a second arm.

As shown by Fig. 18, the shoulder provides three degrees of freedom: linear extension and rotations in elevation and azimuth. All motions are polar-centric, that is, centered on a common point, so that transformation of coordinates can be as simple as possible.

It is planned to map the probings of the finger in a somatic first-stage computer, similar to the visual first-stage computer described in Section 7. The z-axis of that mapping will be depth rather than luminance. Sudden changes in depth will be detected by lateral inhibition, as they are in animals.[8]

17. COMMAND OF THE COMMANDER?

Since both visual and tactile computation can be commanded to respond to some sizes and shapes more strongly than others, we are led to ask: What commands the command computer of an animal? Dr. McCulloch's answer is revealing: "Nothing. You can persuade or cajole the command computer, but you cannot command it." His statement is supported by his diagram in Fig. 2 where the influences upon retic are represented by the many connections to it. "There need to be connections from more than one sense, preferably at least three," he continues. Other influences upon retic are internal, such as the cerebral cortex, where are stored models of the environment, past and present, and future models in the form of goals. These influences may be stronger than

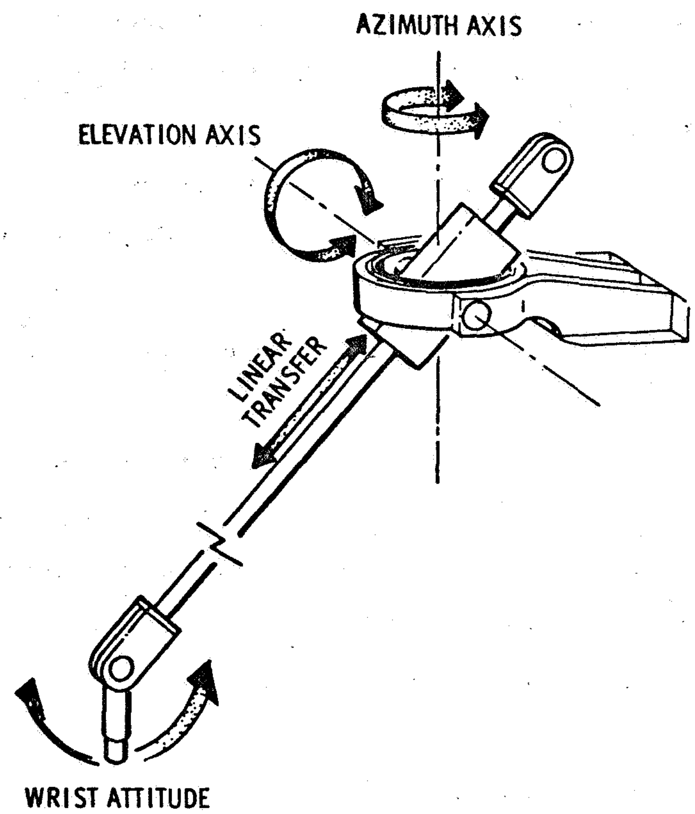


Fig.18 Degrees of freedom of the hand and arm.

external ones. Influences to a model of retic are represented by the S_i of Fig. 17.

Modes, which the retic of a vertebrate animal chooses among, are as follows.[16]

1. sleep
2. eat
3. drink
4. fight
5. flee
6. hunt for prey or fodder
7. explore (or search)
8. urinate
9. defecate
10. groom
11. mate (or sex)
12. give birth or lay egg
13. mother the young (e.g., suckle, hatch, retrieve)
14. build or locate nest
15. innate forms of behavior unique to the species, such as
    - migrate
    - hibernate
    - gnaw
    - hoard

A comparable list for a Mars robot is given in Section 12.

We can imagine how a man's retic selects among his possible modes of behavior. Vision, as we pictured it in Section 7, consists of both the feed-inward of raw sensory data and the feed-outward of signals to adjust filters to match internally-generated images. Touch, hearing and kinesthesia appear to operate in similar fashion. Sensory inputs, therefore, of the kind represented by the S_i of Fig. 13, represent information from external and internal sensors and internal computers.

Consider the actions of a soldier on sentry duty at an advanced post in enemy country. He is expected to hold his filters to match the stimuli he has been taught to expect from the enemy: the shape of his face, the color of his uniform, his manner of fighting, etc.[17] If the soldier hears the snap of a breaking twig, turns toward the sound, and receives from its direction images that match the projections of his stored models, he may classify the sound and the sight as "enemy". What the soldier does next depends upon other models stored in his relational computer. If circumstances appear to favor combat, the sentry may fight (mode 4). If circumstances appear overwhelmingly adverse, the sentry may turn and run (mode 5). Circumstances that cause his command computer to select a mode are the number of the enemy, its armament, etc., as these are recognized by the process of generating projections and comparing these at the filters with incoming images. Thus the enemy can persuade or cajole the soldier's command computer.

Similarly, the Martian environment should be able to persuade or cajole the command computer of a Mars robot, in cooperation with very many fewer relations than are required for a human being, such as a sentry. As far as we know the Mars robot will not have to contend with enemies and, therefore, will not have to fight or flee. Nor will it have to enter any of the other modes which an animal has to enter in order to eat and reproduce.

The ten modes suggested in Section 12 are very much reduced versions of the requirements of an animal. Instead of the elaborate mode of behavior to hunt, are the much simpler ones to advance, retreat, and look-and-feel.

18. STAGES IN THE DEVELOPMENT OF A ROBOT

The system we have just described can be achieved, we believe, by such a process of designing, simulating, and testing as we have described for the development of the separate computers.

A camera-computer chain of the kind described in Sections 7 through 9 can be packaged so as to be mobile. A possible mobile camera-computer chain is shown in Fig. 19. To eliminate the need for its own power supply and transmitter to earth, it is connected to a Mars lander by a cable which it pays out from a spool at its center. To permit it to look from side to side without moving, its camera employs a rotating prismatic mirror which reflects the light of the scene through both left and right optical paths, to photocells which transduce luminances to voltages. The voltage amplitudes, when converted to digital words, then enter shift registers, to move past computing elements in the manner described in Section 9. Vertical scan is attained by turning the drum that holds the camera.

In initial experiments, the visual first-stage computer can be at the far end of the cable in the lander. When a light-weight visual first-stage computer has been completed, it can be placed within the mobile data-gathering element.

The command computer will not be tied in initially, the command being exercised by the earth operator. When the command computer has reached the stage of development where it can be tied in, it can be built into a robot that is physically separate from the lander, in a configuration such as that shown in Fig. 20. This design shows, for test purposes, a stereo

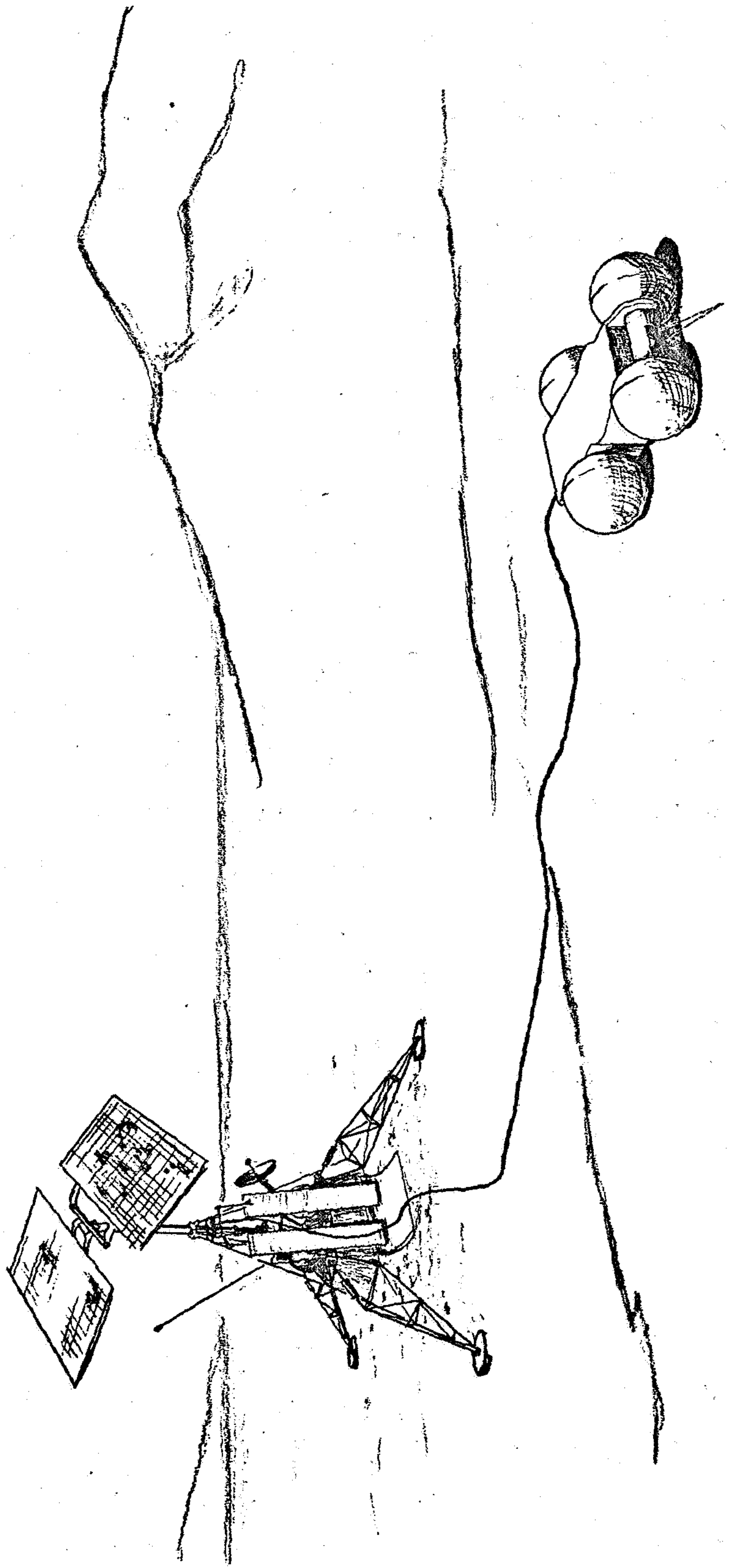


Fig.19 Proposed mobile data-acquiring element for a Mars landing.

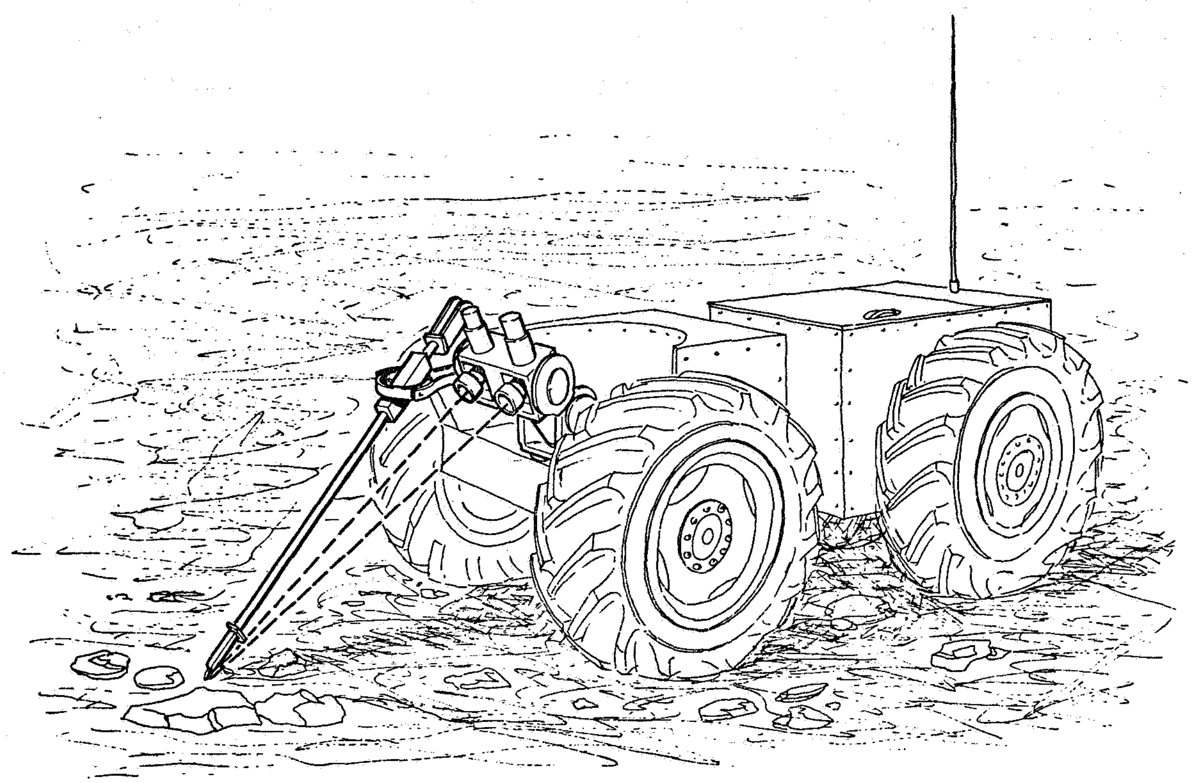


Fig.20 Proposed experimental robot.

television camera mounted in gimbals on a commercially-available tractor. The gimbal mounting of the camera permits it to look forward, sideward, up, down and backward. Beside the camera is the arm described in Section 15. The rubber tires would be replaced for Mars travel by wire-mesh tires.

19. COMPARISON WITH OTHER ROBOT PROJECTS

Perhaps because it is directed toward development of the properties of the computers descrebed in Section 2, ours is the only robot project with a binocular TV camera as input, a visual second-stage computer to employ this binocular input for precision computation of range, and a command computer to receive simultaneous inputs from several senses and decide what class of thing the robot should do.

However, the visual computers and the command computer are still operating sepately while in other robot projects assemblies of computers and external equipment are already operating together. Eye-computer-arm-hand systems are in operation at Project MAC, M.I.T.[18] and the Department of Computer Science at Stanford University[19]. A computer-arm-hand system is in operation at the Department of Mechanical Engineering, M.I.T.[20] An eye-computer-vehicle system is in operation at Stanford Research Institute (SRI)[21]. Out of the efforts of the projects have come list processing languages[22,23] which we may use in simulating a relational computer. Other contributions are speech recognition[19], the kinematics of manipulators under computer control[24], the mapping of the space in which a manipulator operates[25] and recognition of visual contiguity[26].

Since we are all engaged in processing increasingly large volumes of data, reward goes to the one who discovers a method of extracting useful data. At Project MAC and Stanford the reward can take the form of a doctor's degree; at SRI and our project, which are more equipment oriented, the reward is to have either the equipment or a simulation of it work. This reward is also present at the two universities; and SRI and we also do theoretical work.

Aside from the amount of equipment in operation, the greatest difference between our project and the other three is in the way we use the life sciences. All of us use this information since we are all trying to make computers and other hardware do what until now, only animals have done. However, we give primary attention to anatomy and physiology, secondary attention to psychology and invention, while other projects use a reverse emphasis. We make an exhaustive search of the literature on each animal "computer" we investigate and attempt to create, within the constraints of technology, a working model of the "computer". Examples of such literature searches are given in references 28, 4 and 16.

Once we have achieved a working model we proceed to improve it and in doing so may excel nature.

20. CONCLUSION

To some, the work reported here will appear to be slavish imitation of the vertebrate nervous system. It appears to us, on the other hand, to be good engineering practice. When something does what you want it to do and is already miniaturized, why not copy it?

The copy is made only in principle. Visual computers are time-shared elements, past which data is moved by shift registers. The command computer, on the other hand, is designed to be built of modules, each of which forms a larger portion of the whole than do cells in retic. In fact, the modules and the cells are similar only in their mathematics.

To quote Dr. McCulloch, "We use the word 'robot' in two ways. The first, and less important, is as a machine that performs isolated functions of a human being. The second, and more important, is as a description ~~of life~~ applicable to either living things or machines. Such a description is indifferent to whether the system is man-made or grown, whether it is built of hardware or living cells. This is a central theme of cybernetics: to use the same logic and mathematics to describe men and machines. Norbert Wiener looked at control this way. We are looking at both command and control. Thus, in the more important sense, a robot is a prescription for a system that until recently could be achieved only by the growth of living cells but is becoming something we can manufacture."

APPENDIX A

FILTERS

A.1 Scaling

As indicated in Section 7, application of a filter such as W_2 will produce matrices of luminances that are "on scale", except when applied to spots and edges of maximum contrast. By "on scale" we mean on the scale of 0 to 63 that our digital-to-analog converter can convert and our scope can display.

When filter $W_1 \times \frac{1}{8}$ is convolved with the digitized image of Fig. 5a and a histogram plotted of the convolution sums, the result is Fig. 21. The extremes are formed when the center of the submatrix of luminances is 63 and the surround is zero, or vice versa. By lopping off extremely high and low values of the convolution sum, as shown by cross hatching, the scale is contracted. Added to the central luminance of the submatrix in question, the convolution sum enhances contrast. In the rare case where this addition forms a negative sum, this sum is replaced by zero. Filter $W_1 \times \frac{1}{8}$ is used in this example, rather than filter W_2, to demonstrate a symmetrical histogram. The 2 in the center of filter W_2 leads to the unsymmetry.

A.2 Filter to Pass Luminances at the Maximum Useful Spatial Frequency

Luminance contrast is defined [27] as

$$\frac{L_1 - L_2}{L_1} \tag{A1}$$

where L_1 and L_2 are the luminances of the background and object, respectively. Such contrast is determined in animals, not by a simple subtraction and division as here, but by a matrix of interacting filters, as illustrated in Fig. 22. The pattern of response is the net effect of overlapping, and possibly opposed, excitatory and inhibitory influences. This interaction may be expressed by a set of simultaneous equations, one equation for each element; but simultaneous solution of these equations is not possible in a light, portable device.

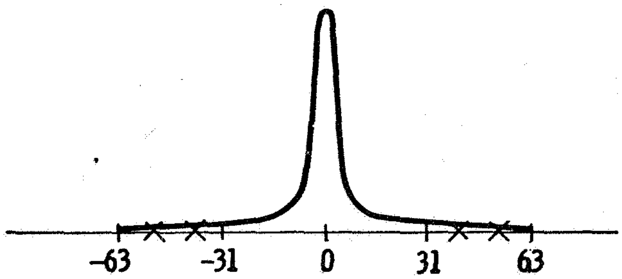


Fig. 21 Histogram of convolution sums formed by filter W_1 with all of the sub-matrices of luminance in a typical scene.

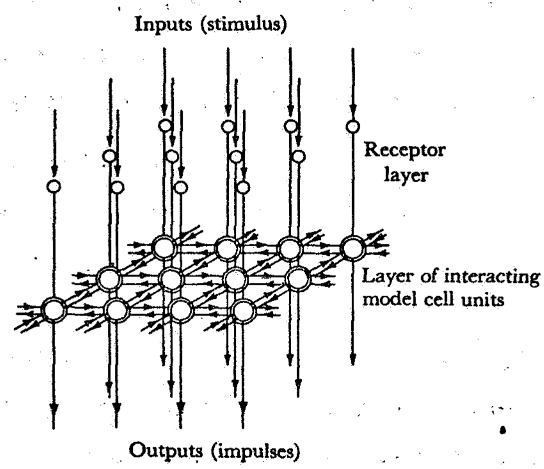


Fig. 22 Schema of a receptive layer and an interacting layer, with arrowheads indicating the direction of conduction of impulses. (Reprinted with permission[28])

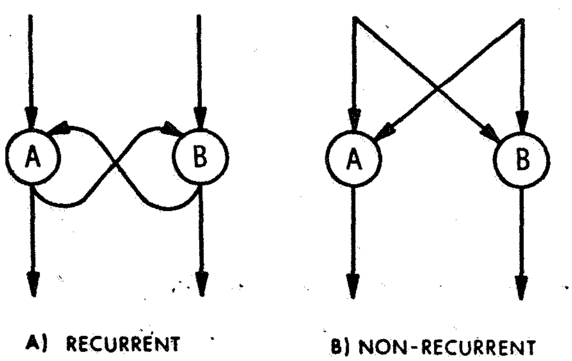


Fig. 23 Interaction of cells A and B.

$$W_3 = \begin{bmatrix} -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 8 & 8 & 8 & -1 & -1 & -1 \\ -1 & -1 & -1 & 8 & 64 & 8 & -1 & -1 & -1 \\ -1 & -1 & -1 & 8 & 8 & 8 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \end{bmatrix} \times \frac{1}{64}$$

Fig. 24 Filter employed to enhance the contrast of Fig. 5. Results are shown in Fig. 6 a It was employed again to enhance the contrast of Fig. 6a with the results shown in Fig. 6b.

Instead, if the interaction of Fig. 22 is represented by Fig. 23 (where the criss-cross lines represent inhibition), and the non-recurrent interaction is selected as the model, a filter can be devised that can be convolved with a matrix of luminances to closely approximate the desired interaction.

As explained in Section 7, application of the filter takes place, in principle, in two steps. In the first step, the image is applied to a filter which (1) passes luminances in the scene that recur at a spatial frequency as high as possible, and (2) rejects luminances that recur because of shot noise and other electronic effects. A filter that does part (1) of this first step is W_1 in Section 7.

Applying digital-signal-processing theory and testing by experiment, Jerome Lerman devised filter W_3 (Fig. 24). This filter achieves parts (1) and (2) of step one, and also performs step two, which is an addition of the result of step one to the original image in a ratio of 1 to 8 (instead of 1 to 1 as in filter W_2). Figure 6a shows the effect of convolving W_3 with the matrix of luminances in Fig. 5. Figure 6b shows the effect of convolving W_3 with the matrix of luminances in Fig. 6a. The effect of contrast enhancement between Fig. 5 and 6a is not so evident here as in the scope displays because photography enhances contrast.

Derivation of W_3, of the filter employed to obtain Fig. 7, and of the means for determining range by comparing left and right views, will be presented in subsequent reports.

Blank

APPENDIX B

OPERATION OF S-RETIC

Each set of lines $\overleftarrow{p}_i$ in Figure 17 (only $\overleftarrow{p}_7$ is marked) indicates the i^{th} module's degree of preference for each of the four possible modes. Thus $\overleftarrow{p}_i$ is a probability vector. The ascending and descending lines out of M_i are branches of $\overleftarrow{p}_i$ which connect to other modules.

As shown in Figure 25, each module has two parts. Each a-part computes from the module's input information the initial guess as to what mode of behavior is best for the robot. The b-part computes an adjusted guess from information received from above, below and from the a-part. The a-part which has five binary input variables and four probabilistic variables, is a non-linear transformation network.

The b-part receives 4-component probability vectors $\overleftarrow{p}_\delta$ from above, $\overleftarrow{p}_\alpha$ from below and $\overleftarrow{p}'$ from the a-part. The components of each vector are the relative probabilities of each of the four possible modes of behavior. The j^{th} component of each p_α, p_δ or p vector is the probability, computed by the module of origin, that the command computer's present γ input signal configuration calls for mode j. The b-part also receives a measure of the cycle-to-cycle difference between γ inputs, called "γ difference" in Figure 25, and a measure, Q, of the strength of convergence on the last mode commanded, (shown being formed in the lower right corner of Fig. 17.

Consensus among the modules is achieved, as illustrated in Fig. 17 by first determining in the step function (s), if the j^{th} component of the probability vector p from each module exceeds 0.5. If it does, a 1 is passed on to the threshold element T_j. There, if 50% or more of the j^{th} component input connections are 1's, mode j is commanded. Note that each element T_j in Fig. 17 receives 10 inputs, although for clarity in the drawing, connections are shown only to T_3. The threshold elements T are part of the executive computer pictured in Fig. 2. Each T receives many inputs, decodes them and executes a mode of behavior.

S-RETIC is described in more detail in Reference 15.

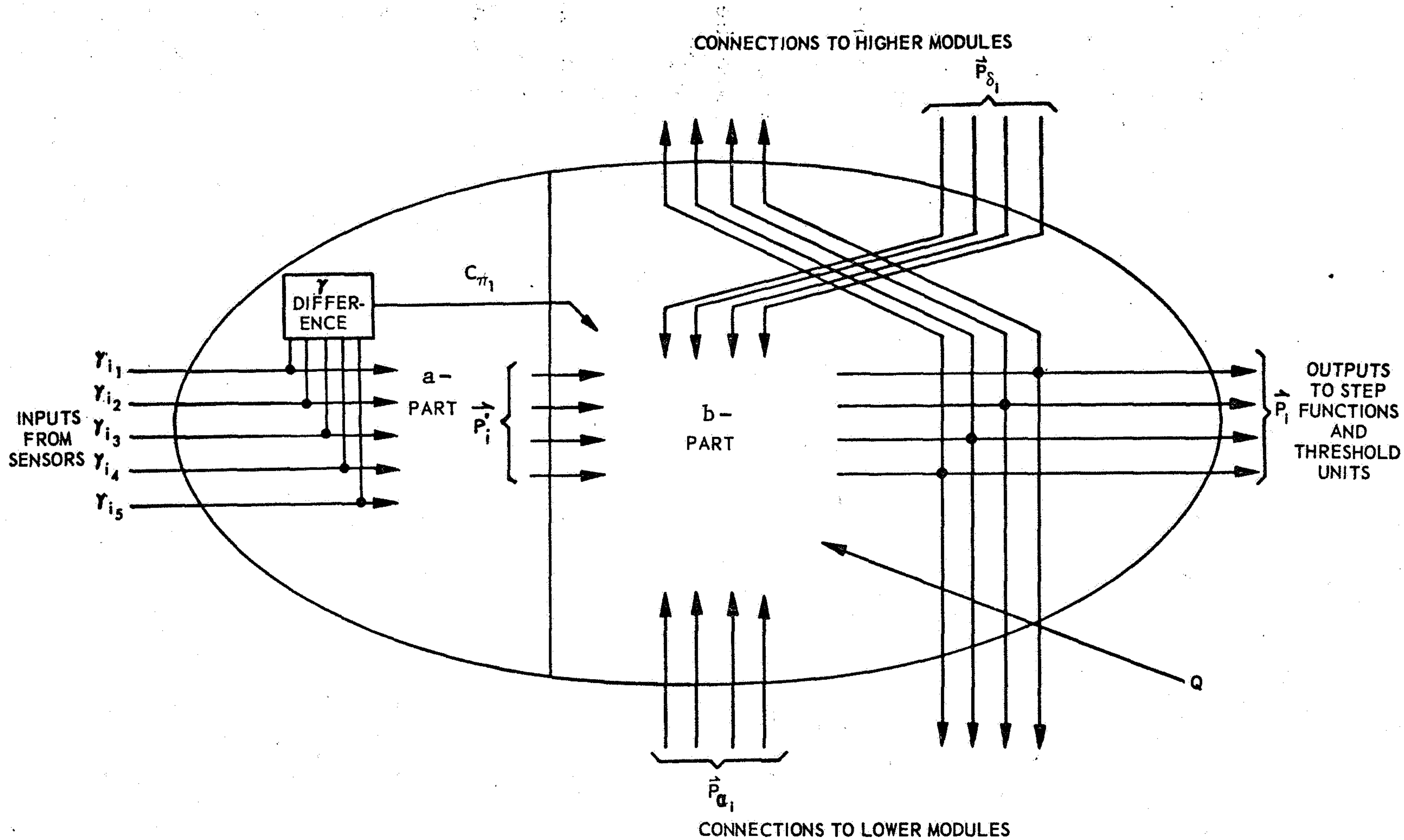


Fig. **25** Input and output connections to parts a and b of a module in S-RETIC.

Note to Section 15

APPENDIX C

OPERATION OF STC-RETIC

C. 1 General

Figure ~~28~~ 26 diagrams the a-part of a module in STC-RETIC. Signals from the senses (ν's), from modules above and below (ω's), from the threshold units that determine the commanded mode (CM) and from the experimenter (RPC and RAC's) enter the short term memory where they are held for two input epochs. STC-RETIC converges on a mode in the same sequence described above for S-RETIC, except that each initial guess is read from a medium-term memory instead of being generated by a transformation network. Each initial guess is read out as a probability vector to be used in the same way as p_i' in S-RETIC (Fig. ~~27~~ 25).

In the conditioning process, the information as to module inputs, module p', the converged-upon mode and whether it was good or bad, is used to modify the probabilities stored in medium term memory.

In either conditioning STC-RETIC, allowing it to habituate or in other ways encouraging it to acquire temporally influenced patterns of behavior, the operator of STC-RETIC not only provides the external conditions (Σ, Fig. 17) to which it will respond, but he takes the place of the parts of a robot that would indicate that the input and response conditions are punishing, neutral, or rewarding. He introduces a negative reinforcement prior to the command (RPC) to indicate that the external conditions are themselves punishing, and a zero RPC to indicate that they are neutral. Finally he introduces four other numbers (RAC's) to indicate the reinforcement effect on STC-RETIC of its command of each possible mode of behavior. An RAC_i is a "reinforcement to be applied after the command if STC-RETIC converges to mode i." It is used in conjunction with short-term-memory information in the module to specify the corresponding modification to the module's medium-term memory. A negative value of RAC_i, called "punishing," is interpreted by STC-RETIC as evidence for not commanding mode i the next time the same ν's and ω's are received. A positive value, called "rewarding," is interpreted oppositely.

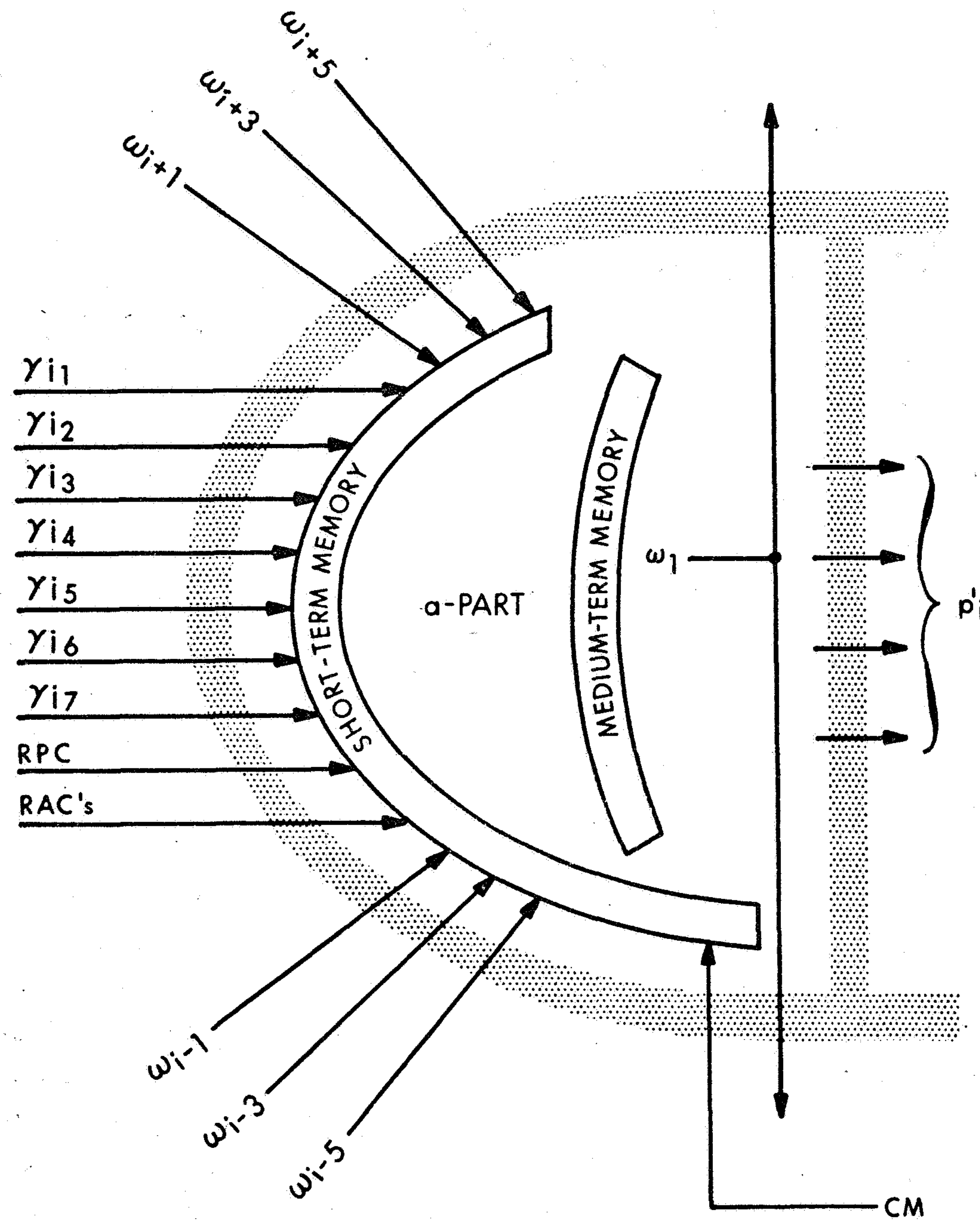


Fig. ~~28~~ 26 The a-part of a module in STC-RETIC.

C. 2 Examples of Pavlovian Conditioning and Habituation in the Robot

Pavlovian conditioning is made possible both by responses "wired into" the medium-term memory of each module of STC-RETIC, and by γ's and ω's acquired during two epochs by the module's short-term memory. For example, let us suppose that a "wired-in" (unconditioned) response is (1) to command retreat when the robot feels a horizontal edge with nothing beyond it (a precipice?), then (2) to be rewarded (positive RAC) for saving the robot. Let us suppose further that each time the robot feels an edge, beyond which it feels nothing, it also sees an edge, beyond which it sees nothing. The more often the robot both feels and sees an edge, beyond which it feels nothing, and then is rewarded for retreating, the more firmly it becomes conditioned to back up at the sight, instead of just the feel, of such an edge.

In STC-RETIC such conditioning is stored in the ~~long~~ medium-term memory of each module, which contains a reduced record of the conditions under which it has made decisions.

An example of habituation in a robot is the following. Suppose a robot travels over the surface of Mars and comes upon territory that is sharply striped in light and dark. When it ceases to behave as though each stripe were a precipice, it will have become habituated to the striped terrain. After a time, the duration of which was pre-specified, it will spontaneously recover the original response so that, if it then moves into territory where such stripes are shadows marking true hazards, it will not be harmed. What have changed in both of these examples are the medium-term memories of modules.

C. 3 Forms of Robot Behavior Due to Development

As explained above, the initial guess of the a-part of each module is made by consulting its long term memory. If there is no entry for the given values of γ and ω, then a flat probability vector (all four components equal) is read out.

Changes in a long term memory, which starts from flat probability vectors, are what we mean by development. STC-RETIC has room in the long term memory of each module for 64 vectors, some of which are flat, others are preprogrammed to non-flat values. Some of the kind of development that is complete in the retic of a mammal at birth may be best achieved by STC-RETIC through experience in its environment.

Other forms of plastic behavior modeled by STC-RETIC are: generalization of and discrimination among the conditions under which a mode response is given, avoidance conditioning and extinction of Pavlovian conditioning.

REFERENCES

1 C M MICHAUX
Handbook of the physical properties of the planet Mars
National Aeronautics & Space Administration 1967 for sale by the Superintendent of Documents U S Government Printing Office Washington D C 20402

2 SPACE SCIENCE BOARD OF THE NATIONAL ACADEMY OF SCIENCES
Planetary exploration 1968-1975
2101 Constitution Avenue N W Washington D C 20418

3 L L SUTRO
Information processing and data compression for exobiology missions
R-545 Instrumentation Laboratory MIT Cambridge Massachusetts 1966

4 R MORENO-DIAZ
An analytical model of the group 2 ganglion cell in the frog's retina
Report E-1858 Instrumentation Laboratory MIT Cambridge Massachusetts 1965

5 W S McCULLOCH
Embodiments of mind
The MIT Press Massachusetts Institute of Technology Cambridge Massachusetts 1965

6 D DENNY-BROWN
The cerebral control of movement
Chapter VI Charles Thomas Publishers 1965

7 R H SELZER
The use of computers to improve biomedical image quality
Proceedings of the Fall Joint Computer Conference 1968

8 G VON BEKESY
Sensory inhibition
Princeton University Press Princeton New Jersey 1967

9 F RATLIFF
Mach bands quantitative studies of neural networks in the retina
Holden-Day Inc. San Francisco 1965

10 L L SUTRO C D SIGWART J D LEAVITT D G TWEED
Instrumentation of camera- computer chains
Report R-636 Instrumentation Laboratory MIT Cambridge Massachusetts 1969

11 R GREGORY
Eye and brain
World University Library McGraw-Hill New York 1966 p 50-60

12 W M BRODEY N LINDGREN
Human enhancement beyond the machine age
Describing the work of Donald McKay IEEE Spectrum February 1968 p 89

13 L L SUTRO
Computer synthesis of images with shades and shadows
Report E-2250 Instrumentation Laboratory MIT Cambridge 1969 (in preparation)

14 W L KILMER W S McCULLOCH
The reticular formation command and control system
Proceedings of the Symposium on Information Processing in the Nervous System, State University of New York at Buffalo October 1968 (in publication)

15 W L KILMER L L SUTRO J BLUM C D SIGWART
Steps toward the automatic command of a rover or spacecraft
Report R-635 Instrumentation Laboratory MIT Cambridge Massachusetts 1969 (in preparation)

16 W L KILMER W S McCULLOCH
The biology of the reticular formation
Michigan State University Division of Engineering Research Report East Lansing Michigan March 1969

17 J D FRANK
The face of the enemy
Psychology Today November 1968 p 24

18 M L MINSKY S A PAPERT
Research on intelligent automata status report II
Project MAC MIT Cambridge Massachusetts 1967

19 J McCARTHY L D EARNEST D R REDDY P J VICENS
A computer with hands eyes and ears
Proceedings of the 1968 Fall Joint Computer Conference

20 W R FERRELL T B SHERIDAN
Supervisory control of remote manipulation
IEEE Spectrum Vol 4 No 10 October 1967 pp 81-88

21 N J NILSSON C A ROSEN B RAPHAEL
G FORSEN L CHAITIN S WAHLSTROM
Application of intelligent automata to reconnaissance
Stanford Research Institute Menlo Park California 1968

22 E C BERKELEY D G BOBROW EDITORS
The programming language LISP
Information International 200 Sixth Street Cambridge Massachusetts 1967

23 L G ROBERTS
Machine perception of three-dimensional solids
Optical and electro-optical information processing MIT Press 1965

24 D L PIEPER
The kinematics of manipulators under computer control
Computer Science Department Stanford University Palo Alto California 1968

25 D E WHITNEY
State space models of remote manipulation tasks
Engineering Projects Laboratory Department of Mechanical Engineering MIT Cambridge Massachusetts 1968

26 A GUZMAN
Some aspects of pattern recognition by computer
MAC-TR-37 Project MAC MIT Cambridge Massachusetts 1967

27 ILLUMINATING ENGINEERING SOCIETY
IES lighting handbook fourth edition
345 E 47th St New York 1966 p 2-6

28 F RATLIFF
Op Cit p 118 (reprinted with permission)